ON RISK PREDICTION

Carl Lönnbark

Abstract

This thesis comprises four papers concerning risk prediction.

Paper [I] suggests a nonlinear and multivariate time series model framework that enables the study of simultaneity in returns and in volatilities, as well as asymmetric effects arising from shocks. Using daily data 2000-2006 for the Baltic state stock exchanges and that of Moscow we find recursive structures with Riga directly depending in returns on Tallinn and Vilnius, and Tallinn on Vilnius. For volatilities both Riga and Vilnius depend on Tallinn. In addition, we find evidence of asymmetric effects of shocks arising in Moscow and in the Baltic states on both returns and volatilities.

Paper [II] argues that the estimation error in Value at Risk predictors gives rise to underestimation of portfolio risk. A simple correction is proposed and in an empirical illustration it is found to be economically relevant.

Paper [III] studies some approximation approaches to computing the Value at Risk and the Expected Shortfall for multiple period asset returns. Based on the result of a simulation experiment we conclude that among the approaches studied the one based on assuming a skewed t distribution for the multiple period returns and that based on simulations were the best. We also found that the uncertainty due to the estimation error can be quite accurately estimated employing the delta method. In an empirical illustration we computed five day Value at Risk's for the S&P 500 index. The approaches performed about equally well.

Paper [IV] argues that the practise used in the valuation of the portfolio is important for the calculation of the Value at Risk. In particular, when liquidating a large portfolio the seller may not face horizontal demand curves. We propose a partially new approach for incorporating this fact in the Value at Risk and in an empirical illustration we compare it to a competing approach. We find substantial differences.

Key words: Finance, Time series, GARCH, Estimation error, Asymmetry, Supply and demand.

Acknowledgements

Tied for the gold medal are my supervisor Kurt Brännäs and my fiancée Sanna Strindberg. Kurt, I am tremendously grateful for all the knowledge you have shared and for all the things you have helped me with during these years. Without exceptions, your door has been open for questions and guidance. First-class supervision! Sanna, thank you for all the support and for being on my side. The time with you has been wonderful so far and I am looking forward to the years to come. I hope that you can find other stimulating tasks for me in the future, now that a stupid machine has replaced me as dishwasher.

I wish to thank Johan Lyhagen and other people mentioned here for reading and commenting on parts of this thesis. Eva Cederblad, Marie Hammarstedt and Kjell-Göran Holmberg, you have provided excellent help on various practical matters. I am also thankful for the financial support from the Jan Wallander and Tom Hedelius Foundation, the Tore Browaldh Foundation and the J C Kempe Memorial Fund.

Camilla Andersson, I truly value the time we have spent together. The many hours at Rött ended up with the matching of our younger siblings. At IKSU we had a lot of fun in the sand. Mikael Lindbäck and Susanne Alldén, thank you for the many games. David Granlund, since the launching of ”Beach 2005/On Economics and Econometrics” in January 2005 we have discussed a lot of things and we have had a lot of fun together. Albina Soultanaeva, we teamed up in the estimation of the monster in the first paper of this thesis. I really enjoyed working with you. Jörgen Hellström, Andrea Mannberg, Ulf Holmberg, Daniel Halvarsson, Catia Cialani, Thomas Broberg, Thomas Jonsson, Lena Birkelöf and the rest of the crew at the Department of Economics, you all contributed to make the years in Umeå a great experience for me.

Henrik Wrethling, I really appreciated the weekly lunches. Iver Baatvik and Lars Qvigstad Sørensen, thank you for making the time in Berkeley enjoyable. Erik-Axel Nord, Johan Jönsson and Tobias Larsson, we got to know each other in Uppsala a long time ago and we have since then been on some really great adventures together.

Finally, I wish to express a big gratitude towards my parents and my brothers with girlfriends for all the support and the many good times. Thank you Maria and Svante, Erik and Caroline (Camilla's younger sister), Gustav and Linnéa and Fredrik!

Solna, April 13, 2009
Carl Lönnbark

This thesis consists of a summary and the following four papers:

[I] Brännäs, K., De Gooijer, J. G., Lönnbark C. and Soultanaeva A. 2008. Simultaneity and Asymmetry of Returns and Volatilities in the Emerging Baltic State Stock Exchanges. Umeå Economic Studies 725 (revised).

[II] Lönnbark, C. 2008. A Corrected Value-at-Risk Predictor. Forthcoming in *Applied Economics Letters*.

[III] Lönnbark, C. 2009. Uncertainty of Multiple Period Risk Predictors. Umeå Economic Studies 768.

[IV] Lönnbark, C., Holmberg, U. and Brännäs, K. 2009. Value at Risk for Large Portfolios. Umeå Economic Studies 769.

Paper II is included with permission from the journal.

1 Introduction

The Oxford Advanced Learner's Dictionary defines risk as "the possibility of something bad happening at some time in the future". There are many different types of risks (even in a financial context) and in this thesis the focus is on a type of risk referred to as market risk, which is the risk of adverse price movements. In the lecture delivered in connection with receiving The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, Robert Engle noted: "The advantage of knowing about risks is that we can change our behavior to avoid them" (Engle, 2004). Of course, as he further notes, we do not wish to avoid them completely. Rather, we take on risks that we think are worthwhile. This trade-off between return and risk is at the heart of financial economics.

Unlike return, risk is something that we never observe directly and knowing about it is almost synonymous to the proper assessment of it (so that it can be managed). Historically, the standard way of measuring risk has been by the variance of asset returns. This is to a large extent due to the huge impact of the modern portfolio theory of Markowitz (1952). However, what risk measure to use is very much context dependent. For discussions of views on risk and reviews of risk measuring, see Granger (2002) and McNeil, Frey, and Embrechts (2005).

The standard way of measuring market risk in the financial industry today was pioneered by J.P. Morgan's RiskMetrics with the Value at Risk (VaR), which was unveiled in 1994 (see J.P. Morgan and Reuters, 1996). In fact, in the second Basel framework (often referred to as Basel II) the Basel Committee on Banking Supervision (Basel hereafter) requires banks and financial institutions to set aside capital buffers in order to meet market risks, which are usually measured by VaR's (Basel, 2006). Thus, accurate VaR's are crucially important for the stability of the financial system and the measure has received a great deal of attention in the literature (e.g., Jorion, 2007). Essentially, the VaR is defined as a potential portfolio loss that most likely will not be exceeded. In statistical terms it is nothing but a quantile of the return distribution.

The attractive feature of the VaR is that it summarizes the properties of the return distribution into an easily interpreted number. However, a major concern for it is that it is silent about the size of the loss when disaster strikes (see Artzner, Delbaen, Eber, and Heath, 1999, for other concerns and a formal discussion of what constitutes a good risk measure). The Expected Shortfall (ES), on the other hand, gives the expected loss, given that the loss exceeds the VaR. The ES is gaining increasing popularity and Yamai and Yoshiba (2005) and others argue in favor of its use. It is the second risk measure studied in this thesis.

Measuring the market risk essentially boils down to making assumptions about the future outcomes of asset returns and it is often closely related to predicting volatility. Obviously, this renders good volatility predictors crucially important and the most popular framework for it is, without doubt, the ARCH and GARCH of Engle (1982) and Bollerslev (1986). In the first paper, we contribute to this field by proposing a model for the joint evolution of the Baltic stock markets.

The assumptions made about the outcomes of future asset returns are associated with uncertainty as well. For example, the assumed risk model may be badly misspecified (see Derman, 1996, for a discussion). Typically, the model is specified up to some parameters and even though the model happens to be a good approximation of reality we still have to estimate those parameters based on historical observations, or by an educated guess. Of course, this is associated with uncertainty too, but this uncertainty tends to be neglected in practise. In papers two and three we give it attention, though. In particular, we find that the uncertainty due to the estimation error may be substantial and that it has an effect on the interpretation of the VaR.

Another issue that may arise is that of how to best predict the risk on a particular horizon. Most naturally, one could specify a model for the relevant horizon directly. However, as noted above, a risk model is associated with an estimation error and this is directly related to the size of the dataset. Hence, it may be the case that the data at hand does not suffice for a reliable prediction. The alternative is then to specify a model for a higher frequency and to use this model to get an indirect prediction. This is easier said than done and in the third paper we consider some approaches for this indirect prediction problem.

Lastly, consider a large portfolio that contains many shares of an asset. A conventional assumption made in the literature is that the entire position can be sold at the same price. This can be a quite misleading valuation approach, since for a large enough position the seller (buyer) of an asset does not face a horizontal demand (supply) curve. In the fourth paper we incorporate this fact in the VaR.

In what follows, the issues indicated above are further developed and the contributions of this thesis are related to the existing literature. First of all, the VaR and the ES are formally introduced.

2 VaR and ES

We wish to quantify the risk in a portfolio of financial assets between the times T and $T+k$ and to introduce the VaR and the ES we denote by $\mathbf{w} = (w_1, ..., w_M)'$ the time invariant vector of portfolio weights. The log-return (return) between T and $T+k$ for the portfolio is approximately $\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'(\mathbf{y}_{T+1} + ... + \mathbf{y}_{T+k})$, where $\mathbf{y}_{T+l} = (y_{1,T+l}, ..., y_{M,T+l})'$, $l = 1, ..., k$, is a M-dimensional vector of one-period returns. The conditional VaR for the period T to $T+k$ satisfies

$$\Pr\left(\mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha} | \mathcal{F}_T\right) = \alpha, \tag{1}$$

where $\mathcal{F}_T$ is the information available at T and α is a small probability. The associated conditional ES is defined as

$$ES_{T,k}^{1-\alpha} = -E_T\left(\mathbf{w}'\mathbf{Y}_{T,k} \mid \mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha}\right), \tag{2}$$

where $E_T(\cdot)$ is shorthand for expectation conditional on $\mathcal{F}_T$. The minus signs in (1) and (2) stem from the convention of reporting the VaR and the ES as positive numbers.

The $\mathcal{F}_T$ typically contains past asset returns and the goal is to use this information in the best possible way to compute predictors $\widehat{VaR}_{T,k}^{1-\alpha}$ and $\widehat{ES}_{T,k}^{1-\alpha}$. From (1) it is obvious that the VaR is a quantile of the return distribution. Thus, predicting the VaR essentially amounts to employing statistical techniques for quantile estimation. These have been around for a long time and approaches to computing the measures range from non-parametric to fully parametric ones, with lots of hybrids in between. For recent surveys of existing approaches, see Jorion (2007) and McNeil et al. (2005). See also Kuester, Mittnik, and Paolella (2006), for a comparison of some popular alternatives. For example, assuming that the information at hand is a sample of identically and independently distributed (iid) returns a straightforward predictor of the VaR is a suitable order statistic. This approach is referred to as historical simulation in the financial industry.

In this thesis we consider parametric approaches and we assume that the vector process, $\mathbf{y}_t$, of the assets returns started in the infinite past and that it is generated in discrete time up through, at least, $T+k$ by

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{u}_t, \quad \mathbf{u}_t = \mathbf{H}_t^* \boldsymbol{\varepsilon}_t. \tag{3}$$

Conditional on the information available at $t-1$, $\boldsymbol{\varepsilon}_t$ has mean $\mathbf{0}$ and the identity matrix, $\mathbf{I}$, as its variance-covariance matrix. Then, $\boldsymbol{\mu}_t$ is the conditional mean of $\mathbf{y}_t$, whereas $\mathbf{H}_t = \mathbf{H}_t^* \mathbf{H}_t^{*\prime}$ is the conditional variance-covariance matrix. In the next section we discuss specifications of $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$.

3 GARCH

In line with the hypothesis of efficient markets, asset prices are widely taken to be random walks (Gourieroux and Jasiak, 2001) and the effort in terms of modeling is often made on the variance part of (3). Thus, for the conditional mean function various ARMA specifications are routinely adopted (e.g., McAleer and Da Veiga, 2008). An interesting alternative is the use of the asymmetric moving average model of Wecker (1981) in Brännäs and De Gooijer (2004). Brännäs and Soultanaeva (2006) later extended it to include explanatory variables.

The most popular framework when it comes to the modeling of the conditional variance is the GARCH. Since Engle's seminal paper, the ARCH-literature has exploded with extensions of the basic model; adapting it to different stylized facts of financial asset returns (see Cont, 2001, for an account of stylized facts). In fact, the GARCH models were originally developed to cope with the stylized fact of volatility clustering. For a survey on GARCH models and other volatility predictors, e.g., models of stochastic volatility, see Andersen, Bollerslev, Christoffersen, and Diebold (2006).

The workhorse specification (cf. Hansen and Lunde, 2005) in univariate situations, i.e. $\mathbf{H}_t = h_t$, is the GARCH(1,1) model

$$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}, \tag{4}$$

where $u_t = y_t - \mu_t = \sqrt{h_t}\varepsilon_t$, i.e. the one-period ahead prediction error. A stylized fact that has proved highly relevant empirically is the so-called leverage effect, i.e. that negative returns are followed by higher volatility than positive ones. The leverage effect was first acknowledged by Black (1976) and it has been incorporated in the GARCH framework by Glosten, Jagannathan, and Runkle (1993), Nelson (1991) and many others. The model of the former appears to be the most popular one in empirical work and it extends (4) by the term $\gamma \min(0, u_t)u_t$, thus allowing positive and negative shocks to affect future volatility asymmetrically.

In financial contexts we usually deal with portfolios, i.e. we are interested in the joint evolution of several assets or markets. Consequently, multivariate models with variance specifications of the GARCH-type have been developed (see Bauwens, Laurent, and Rombouts, 2006, for a survey). The important feature that multivariate models wish to capture is that of how shocks transmit across assets and markets (e.g., Karolyi, 1995; Bonfiglioli and Favero, 2005). Understanding the nature of this transmission is of great practical interest, as it may have consequences for, e.g., risk management decisions (Fleming, Kirby, and Ostdiek, 1998).

The intra-day literature suggests that information processing is very fast (e.g., Engle and Russell, 1998). Hence, for models specified on a (say) daily frequency it may be important to incorporate simultaneous effects. Indeed, structural VAR models have quite recently been employed to study the joint behavior and contemporaneous interaction among asset returns (e.g., Rigobon and Sack, 2003; De Wet, 2006; Lee, 2006). Obviously, and perhaps more interestingly, there is also reason to expect simultaneous effects in volatilities. Gannon and Choi (1998) and Gannon (2004, 2005) have addressed this question in terms of realized volatilities, i.e. squared returns. However, in a multivariate GARCH context it seems natural to allow for simultaneity in conditional variances. In the first paper of this thesis we propose the, to our knowledge, first model with this particular feature. We apply the model and study the joint evolution of returns and volatilities of the stock exchanges in the Baltic cities Riga, Tallinn and Vilnius.

4 Estimation error

The task of computing the VaR and the ES is predictive in nature and it is clearly subject to uncertainty. Hendry (2000) discusses various error sources in prediction. Here, we focus on the error that is due to the fact that the parameters of the hypothesized model of the data-generating process are unknown and must be estimated. The additional uncertainty from this error source should be of concern to risk managers. Surprisingly little work has been done on it, though. In fact, Lan, Hu, and Johnson (2007) report that the research on the uncertainty of VaR predictors only amounts to about 2.5 percent of the VaR literature. Jorion (1996) was the first to attempt to formally quantify it, but following his paper this research area appears to have rested for some time and regained interest quite recently. For example, Christoffersen and Gonçalves (2005) used resampling techniques to study the uncertainty of VaR and ES predictors in a GARCH framework. The obvious disadvantage of their method is that it is time consuming since it amounts to repeated estimation of a possibly complicated model. Analytical expressions (when sufficiently accurate) to quantify the uncertainty are obviously preferred. For this purpose Chan, Deng, Peng, and Xia (2007) and others consider the conventional delta method, which is done here as well.

In what follows, we will take as given a consistent and asymptotically normally distributed estimator, $\hat{\boldsymbol{\theta}}$, that is centered at the true parameter vector. When $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$ in (3) are correctly specified, one such estimator is the traditional (conditional) maximum likelihood (ML) estimator with a normality assumption on $\boldsymbol{\varepsilon}_t$. This assumption does not fare well with the stylized fact of conditionally leptokurtic and sometimes conditionally skewed asset return distributions. However, as shown by Weiss (1984, 1986) and Bollerslev and Wooldridge (1992), the estimator remains consistent and asymptotically normal even if the distribution of $\boldsymbol{\varepsilon}_t$ is non-normal and it is then known as the Quasi-Maximum Likelihood (QML) estimator. Of course, ML estimation has been employed with other distributional assumptions as well. For example, Bollerslev (1987) considers the Student's t distribution.

Early attempts (e.g., Schmidt, 1974) to quantify the effect on prediction of errors in parameters relied on the asymptotic distribution of the parameter estimator, assumed to be independent of the conditioning information. In the notation set out above, the predictors are functions of $\mathcal{F}_T$ both directly and indirectly through $\hat{\boldsymbol{\theta}}$. Denote this (continuous) function by $u[\mathcal{F}_T, \hat{\boldsymbol{\theta}}(\mathcal{F}_T)]$. The approach then amounts to conditioning the first argument of $u(\cdot)$ on a realization of $\mathcal{F}_T$ and viewing randomness to arise through the random $\mathcal{F}_T$ in the second argument. This approach now appears to be the conventional (see Kaibila and He, 2004, for a recent discussion). Indeed, Hansen (2006) takes this route and shows asymptotic normality for $\widehat{VaR}_{T,1}^{1-\alpha}$.

Now, the question a practitioner naturally poses is how uncertainty in the VaR affects risk management, i.e. does it in some way change what value to report. Indeed, Tsay (2005, ch. 7) points out that the VaR should be computed using the predictive distribution of returns, and it should take into account the parameter uncertainty in a properly specified model. In the second paper we accept this challenge and demonstrate a way of incorporating the estimation error in a VaR predictor. The key insight is that, in practise, we do not use the VaR that satisfies (1), i.e. the true VaR. Instead, we use a random predictor of it and the relevant probability is $\Pr(\mathbf{w}'\mathbf{Y}_{T,k} \leq -\widehat{VaR}_{T,k}^{1-\alpha} | \mathcal{F}_T)$. Clearly,

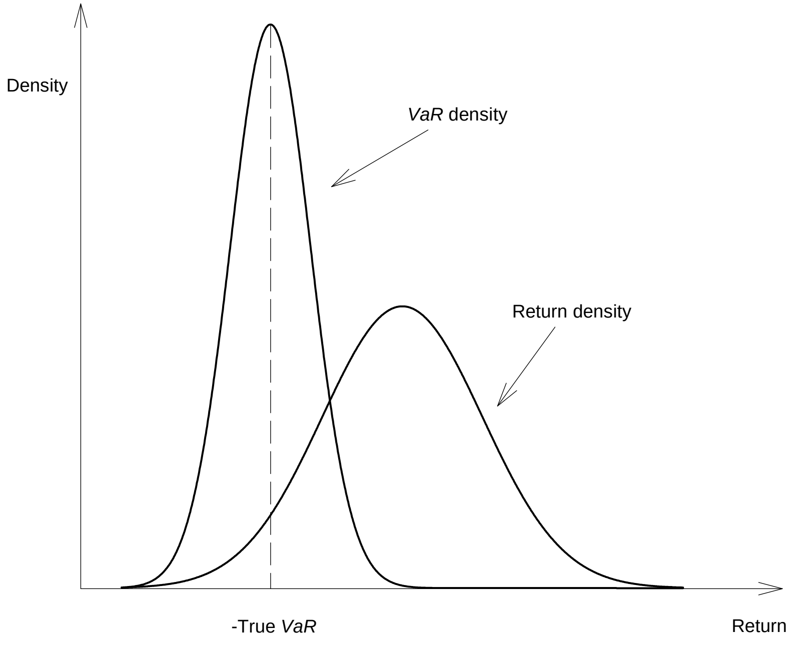

Figure 1: VaR density and return density refers to the conditional densities of the VaR predictor and the return, respectively.

this probability is not necessarily equal to α. In Figure 1 we depict the situation.

Related discussions appear in Schaller (2002) and Escanciano and Olmo (2008). The latter is given in a back-testing[1] context, though. We emphasize that the situation is not bias in the conventional sense, i.e. that the expected value of the VaR predictor is different from the true value. For studies of conventional bias, see Bao and Ullah (2004), Gomes and Pestana (2007) and Hartz, Mittnik, and Paolella (2006).

[1]Back-testing is the blanket term for statistical techniques of VaR predictor validation (e.g., Campbell, 2005).

5 Horizon

It is sometimes of interest to measure the risk on horizons longer than (say) one day. An important example when this is the case is for the market risk charge in Basel II, that is based on an horizon of 10 trading days. It is then natural to specify a risk model for the relevant horizon directly. Indeed, this is the recommendation put forth by Diebold, Hickman, Inoue, and Schuermann (1997). However, as noted above VaR and ES predictors are subject to an estimation error, which is directly related to the sample size. Thus, it may be the case that the available sample size is not large enough for reliable predictions. The alternative is then to specify a model for a higher frequency and iterate on this model to obtain predictions for the relevant horizon. This corresponds to the case $k > 1$ in (1) and (2) and the properties of the multiple period returns are thus of interest.

Now, assume that the one-period portfolio return is normally and independently distributed (nid) with zero mean and variance σ^2. Then, the k-period return is nid with zero mean and variance $k\sigma^2$. In this case the task of computing the VaR and the ES for the multiple period returns is trivial and they are simply obtained by scaling the one-period measures by a factor $\sqrt{k}$. This is the so-called Root-k approach and it is allowed in Basel II. However, it is safe to say that asset returns are not normally distributed and certainly not independent in time and it is well known that this approach may give erroneous VaR's (see Brummelhuis and Guégan, 2005; Brummelhuis and Kaufmann, 2007, for discussions). Thus, alternative approaches are called for and this is the focus of the third paper in this thesis.

Suppose now that the risk manager wants to assess the k-period risk in the portfolio and decides to employ the iterating approach within the GARCH framework. A problem that arises is then that the properties of the multiple period return distribution may not follow easily from the one-period model. For example, even though the multiple period conditional variance implied from a one-period GARCH model with normal innovations is tractable, less so is the distribution of the corresponding innovation (e.g., Boudoukh, Richardson, and Whitelaw, 1997). Two ways to go about it are to use simulations (e.g., Christoffersen, 2003) or to consider some other (than the Root-k) analytical approximation.

To explain the simulation based approach, we first assume that the model (3) have been estimated based on observations available up through T. Based on some assumption on the distribution of $\boldsymbol{\varepsilon}_{T+l}$, $l = 1, ..., k$, we then simulate future k-period portfolio returns and compute the VaR and the ES as empirical counterparts.

Refinements of this basic setup include for example the use of kernel functions for increased efficiency (Scaillet, 2004; Chen and Tang, 2005; Chen, 2008) and extreme value theory (McNeil and Frey, 2000). As for the distributional assumption it is of course natural to maintain the one used for estimation in a maximum likelihood framework. However, an approach that has gained popularity is the so-called filtered historical simulation, that was proposed in a univariate context by Barone-Adesi, Bourgoin, and Giannopoulos (1998), Diebold, Schuermann, and Stroughair (1998) and Hull and White (1998). It involves estimating (3) by QML and the distribution of $\boldsymbol{\varepsilon}_{T+l}$, $l = 1, ..., k$, is approximated by the empirical distribution of the standardized residuals (see also Christoffersen, 2009, for a multivariate extension).

For the analytical approximations we assume that the conditional mean, $\boldsymbol{\mu}_{T,k}$, and the conditional variance-covariance matrix, $\mathbf{H}_{T,k}$, of $\mathbf{Y}_{T,k}$ are tractable, and that the k-period portfolio return admits the scale-location representation

$$\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'\boldsymbol{\mu}_{T,k} + \varepsilon_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{5}$$

where $\varepsilon_{T,k}$ has zero mean, unit variance, and (intractable) conditional density function $g_{T,k}(\cdot)$. The problem then boils down to that of finding a suitable approximation for $g_{T,k}(\cdot)$, and in the third paper we study two alternatives for this. The first approach was originally proposed by Wong and So (2003, 2007), and it involves a fully parametric assumption. The second approach employs a Gram-Charlier expansion (e.g., Baillie and Bollerslev, 1992; Jondeau and Rockinger, 2001).

Alternative approaches based on (5) include Fan and Gu (2003) and

Cotter (2007). The former employ non-parametric techniques on the standardized residuals, while the latter scales the one-period VaR relying on an extreme value theory argument. Taylor (1999, 2000) propose a regression quantile approach that may be viewed as a combination of the direct and the iterating approach.

6 Valuation

In the computation of the VaR and the ES it is often assumed that the assets in the portfolio may be traded at mid-prices[2]. For small positions and with tight spreads[3] this may work fine, but it is not a fair valuation approach in general. For example, trading typically does not occur at mid-prices, but at the best bid and ask prices. Consequently, early adjustments to the VaR focused on incorporating adverse movements in the spread (e.g., Bangia, Diebold, Schuermann, and Stroughair, 1999). However, the seller (buyer) of large enough positions does not face horizontal demand (supply) curves. Hence, the liquidation of a position may give rise to an adverse price impact that goes beyond the spread. The question of how to incorporate this fact in the VaR is a relatively old one and several approaches have been proposed (see Ernst, Stange, and Kaserer, 2009; Stange and Kaserer, 2009, for overviews). In particular, the way to go about it depends on what type of market the asset in question is traded on (see Gourieroux and Jasiak, 2001, ch. 14, for an account of the characteristics of quote-driven and order-driven markets). On quote-driven markets one or several market makers set a bid and an ask price, and the additional information available is essentially transaction data. On order-driven markets (with visible limit order books), on the other hand, it is possible to infer the actual price per share that would be obtained upon immediate liquidation. Indeed, Giot and Grammig (2006) use this information and propose an adjusted VaR. This approach appears to us as the most sound of the existing ones, but, of course, it is of limited applicability on quote-driven mar-

[2] The mid-price is the average of the best bid price and the best ask price.

[3] The spread is the difference between the best bid price and the best ask price.

kets. In the fourth paper we build on the approach in Giot and Grammig (2006) and we give our views on how to adjust the VaR with limit order book data at hand.

The discussion above is viewed as a source of liquidity risk in the literature and it is very relevant in practise (e.g., Malz, 2003). Liquidity risk has received interest from the regulatory side as well (see Basel, 2008).

7 Summary of the papers

Paper [I]: Simultaneity and Asymmetry of Returns and Volatilities in the Emerging Baltic State Stock Exchanges

The paper suggests a nonlinear time series model framework that enables the study of simultaneity in returns and volatilities, as well as asymmetric effects arising from shocks. Using daily data 2000-2006 we study the joint evolution of returns and volatilities in the indices of the Baltic state stock exchanges. Shocks from the Moscow stock exchange enters the model through exogenous explanatory variables. As a motivation for the study we take the potential presence of cross market linkages and information spillovers in international investment and risk management decisions. It appears reasonable to expect that these features are of importance for the markets under study, as they are geographically close and share other common features.

The estimation results indicate recursive structures with Riga directly depending in returns on Tallinn and Vilnius, and Tallinn on Vilnius. For volatilities both Riga and Vilnius depend on Tallinn. In addition, we find evidence of asymmetric effects on both returns and volatilities of shocks arising in Moscow and in the Baltic states.

The practical use of the model is outlined and studied. In particular, we compare portfolio allocations and VaR's obtained from our model to those implied by univariate models. We find substantial differences.

Paper [II]: A Corrected Value-at-Risk Predictor

We argue that the additional uncertainty due to the estimation error matters for the interpretation of VaR predictors. In particular, we demonstrate that reported VaR's may be too small, in the sense that the probability that a portfolio loss exceeds the predicted VaR is higher than desired. A simple way of correcting a VaR predictor to give the correct interpretation is proposed. The approach relies on the so-called delta method of computing the approximative variance of the sampling distribution of the VaR predictor. In numerical and empirical illustrations we verify statistical and economic significance, respectively.

Paper [III]: Uncertainty of Multiple Period Risk Predictors

The focus of this paper is on predicting the VaR and the ES for multiple period asset returns. In general, the properties of the conditional distribution of multiple period returns do not follow easily from the one-period data generating process. This renders computation of the VaR and the ES for multiple period returns a non-trivial task and we consider some approaches to approximating these measures. The first one is the Root-k approach that simply scales the one-period measures by the square root of the number of periods. The second one targets the measures by means of simulations. The third and the fourth approaches derive the measures from analytical approximations to the conditional density of the multiple period returns. We consider a skewed t distribution and a Gram-Charlier expansion. In addition, we view the additional uncertainty due to the estimation error as important and keep it an integral part of the paper. In particular, we study the usefulness of the so-called delta method.

Based on the result of a simulation experiment we conclude that among the approaches studied the one based on assuming a skewed t distribution for the multiple period returns and that based on simulations were the best. The predictors based on the Gram-Charlier expansion and the Root-k approximation showed positive and negative bias, respectively. Except for the Root-k approach and in some cases for the

Gram-Charlier approach we found that the uncertainty due to the estimation error can be quite accurately estimated employing the delta method.

In an empirical illustration all predictors performed about equally well in predicting five day VaR's for the S&P 500 index.

Paper [IV]: Value at Risk for Large Portfolios

In this paper we address the question of how to properly assess the risk in large positions of financial assets. We argue that the practise used in the valuation of the portfolio is of importance for the calculation of the VaR. Commonly, it is assumed that the entire position can be sold at the market price (or mid-price), though one realizes that this can be a quite misleading valuation approach. The reason is that for a large enough position the seller of an asset does not face a horizontal demand curve. Instead, we argue, a portfolio should be valued at the actual prices that would be obtained upon immediate liquidation. Based on a model for the dynamics of the limit order book we propose a partially new approach for incorporating the argument in an intra-day VaR. In an empirical illustration we found substantial differences between our VaR and a competing alternative.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P., Diebold, F. X., 2006. Volatility and correlation forecasting. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 1. North Holland, Amsterdam.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Baillie, R. T., Bollerslev, T., 1992. Prediction in dynamic models with time-dependent conditional variances. Journal of Econometrics 52, 91–113.

Bangia, A., Diebold, F. X., Schuermann, T., Stroughair, J. D., 1999. Modeling liquidity risk, with implications for traditional market risk measurement and management. Working paper 99-06, Wharton Financial Institutions Center.

Bao, Y., Ullah, A., 2004. Bias of a Value-at-Risk estimator. Finance Research Letters 1, 241–249.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don't look back. Risk 11, 100–104.

Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision, 2008. Principles for Sound Liquidity Risk Management and Supervision. Bank for International Settlements, Basel.

Bauwens, L., Laurent, S., Rombouts, J. V. K., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics 21, 79–109.

Black, F., 1976. Studies in stock price volatility changes. In: 1976 Proceedings of the Business and Economic Statistics Section. American Statistical Association, pp. 177–181.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69, 542–547.

Bollerslev, T., Wooldridge, J. M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Bonfiglioli, A., Favero, C. A., 2005. Explaining co-movements between stock markets: The case of US and Germany. Journal of International Money and Finance 24, 1299–1316.

Boudoukh, J., Richardson, M., Whitelaw, R. F., 1997. Investigation of a class of volatility estimators. Journal of Derivatives 4, 63–71.

Brännäs, K., De Gooijer, J. G., 2004. Asymmetries in conditional mean and variance: Modelling stock returns by asMA-asQGARCH. Journal of Forecasting 23, 155–171.

Brännäs, K., Soultanaeva, A., 2006. Influence of news in Moscow and New York on returns and risks on Baltic state stock indices. Umeå Economic Studies 696, Department of Economics, Umeå University.

Brummelhuis, R., Guégan, D., 2005. Multiperiod conditional distribution functions for conditionally normal GARCH(1,1) models. Journal of Applied Probability 42, 426–445.

Brummelhuis, R., Kaufmann, R., 2007. Time scaling of VAR in GARCH(1,1) and AR(1)-GARCH(1,1) processes. Journal of Risk 9, 39–94.

Campbell, S. D., 2005. A review of backtesting and backtesting procedures. Finance and Economics Discussion Series 2005-21, Divions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C.

Chan, N. H., Deng, S.-J., Peng, L., Xia, Z., 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics 137, 556–576.

Chen, S. X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6, 87–107.

Chen, S. X., Tang, C. Y., 2005. Nonparametric inference of value-at-risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

Christoffersen, P., 2003. Elements of Financial Risk Management. Academic Press, San Diego.

Christoffersen, P., 2009. Value-at-risk models. In: Andersen, T. G., Davis, R. A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin Heidelberg, forthcoming.

Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. Journal of Risk 7, 1–28.

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1, 223–236.

Cotter, J., 2007. Varying the VaR for unconditional and conditional environments. Journal of International Money and Finance 26, 1338–1354.

De Wet, W. A., 2006. A structural GARCH model: An application on South African data. Economic Modelling 23, 775–791.

Derman, E., 1996. Model risk. Quantitative strategies research notes, Goldman Sachs.

Diebold, F. X., Hickman, A., Inoue, A., Schuermann, T., 1997. Converting 1-day volatility to h-day volatility: Scaling by Root-h is worse than you think. Working paper 97-34, Wharton Financial Institutions Center.

Diebold, F. X., Schuermann, T., Stroughair, J., 1998. Pitfalls and opportunities in the use of extreme value theory in risk management. In: Refenes, A.-P. N., Moody, J. D., Burgess, A. N. (Eds.), Advances in Computational Finance. Kluwer Academic Publishers, Amsterdam, pp. 3–12.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F., 2004. Risk and volatility: Econometric models and financial practise. The American Economic Review 94, 405–420.

Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Ernst, C., Stange, S., Kaserer, C., 2009. Measuring market liquidity risk - which model works best? CEFS Working paper series 1, Center for Entrepreneurial and Financial Studies, Technische Universität München.

Escanciano, J. C., Olmo, J., 2008. Backtesting parametric value-at-risk with estimation risk. Journal of Business and Economic Statistics, forthcoming.

Fan, J., Gu, J., 2003. Semiparametric estimation of value at risk. Econometrics Journal 6, 261–290.

Fleming, J., Kirby, C., Ostdiek, B., 1998. Information and volatility linkages in the stock, bond, and money markets. Journal of Financial Economics 49, 111–137.

Gannon, G., 2004. Simultaneous volatility transmission and spillover effects. School Working Paper Series 2004/10, School of Accounting, Economics and Finance, Deakin University.

Gannon, G., 2005. Simultaneous volatility transmissions and spillover effects: U.S. and Hong Kong stock and futures markets. International Review of Financial Analysis 14, 326–336.

Gannon, G., Choi, D., 1998. Structural models: Intra/inter-day volatility transmission and spillover persistence of the HSI, HSIF and S&P500 futures. International Review of Financial Analysis 7, 19–36.

Giot, P., Grammig, J., 2006. How large is liquidity risk in an automated auction market. Empirical Economics 30, 867–887.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gomes, M. I., Pestana, D., 2007. A sturdy reduced-bias extreme quantile (VaR) estimator. Journal of the American Statistical Association 102, 280–292.

Gourieroux, C., Jasiak, J., 2001. Financial Econometrics. Princeton University Press, Princeton, New Jersey.

Granger, C. W. J., 2002. Some comments on risk. Journal of Applied Econometrics 17, 447–456.

Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. Journal of Econometrics 135, 377–398.

Hansen, P. R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873–889.

Hartz, C., Mittnik, S., Paolella, M., 2006. Accurate value-at-risk forecasting based on the normal-GARCH model. Computational Statistics & Data Analysis 51, 2295–2312.

Hendry, D., 2000. A general forecast-error taxonomy. Econometric Society World Congress 2000 Contributed Papers 0608, Econometric Society.

Hull, J., White, A., 1998. Incorporating volatility updating into the historical simulation method for value at risk. Journal of Risk 1, 5–19.

Jondeau, E., Rockinger, M., 2001. Gram-Charlier densities. Journal of Economic Dynamics and Control 25, 1457–1483.

Jorion, P., 1996. Risk^2: Measuring the Risk in Value at Risk. Financial Analysts Journal 52, 47–56.

Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill, New York.

J.P. Morgan and Reuters, 1996. RiskMetrics - Technical Document, 4th Edition. J.P. Morgan and Reuters, New York.

Kaibila, P., He, Z., 2004. The adjustment of prediction intervals to account for errors in parameter estimation. Journal of Time Series Analysis 25, 351–358.

Karolyi, A., 1995. A multivariate GARCH model of international transmissions of stock returns and volatility: The case of the United States and Canada. Journal of Business and Economic Statistics 13, 11–25.

Kuester, K., Mittnik, S., Paolella, M. S., 2006. Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4, 53–89.

Lan, Y., Hu, Z. M., Johnson, J., 2007. Can value-at-risk be made more valuable? Resampling techniques for risk uncertainty. Working paper, Financial Studies Discipline, UWA Business School, The University of Western Australia.

Lee, K. Y., 2006. The contemporaneous interactions between the U.S., Japan, and Hong Kong stock markets. Economics Letters 90, 21–27.

Malz, A., 2003. Liquidity risk: Current research and practise. RiskMetrics Journal 4, 35–72.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

McAleer, M., Da Veiga, B., 2008. Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model. Journal of Forecasting 27, 1–19.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271–300.

McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press, Princeton, New Jersey.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–70.

Rigobon, R., Sack, B., 2003. Spillovers across U.S. financial markets. Working Paper 9640, National Bureau of Economic Research.

Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115–129.

Schaller, P., 2002. Uncertainty of parameter estimates in VAR calculations. Working paper, Bank Austria.

Schmidt, P., 1974. The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. Econometrica 42, 303–309.

Stange, S., Kaserer, C., 2009. Market liquidity risk - an overview. CEFS Working paper series 4, Center for Entrepreneurial and Financial Studies, Technische Universität München.

Taylor, J. W., 1999. A quantile regression approach to estimating the distribution of multiperiod returns. Journal of Derivatives 7, 64–78.

Taylor, J. W., 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 19, 299–311.

Tsay, R. S., 2005. Analysis of Financial Time Series, 2nd Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

Wecker, W. E., 1981. Asymmetric time series. Journal of the American Statistical Association 76, 16–21.

Weiss, A. A., 1984. ARMA models with ARCH errors. Journal of Time Series Analysis 5, 129–143.

Weiss, A. A., 1986. Asymptotic theory for ARCH models: Estimation and testing. Econometric Theory 2, 107–131.

Wong, C.-M., So, M. K. P., 2003. On conditional moments of GARCH models, with applications to multiple period value at risk estimation. Statistica Sinica 13, 1015–1044.

Wong, C.-M., So, M. K. P., 2007. On the estimation of multiple period expected shortfall and median shortfall using GARCH models. Working paper, Hong Kong University of Science and Technology.

Yamai, Y., Yoshiba, T., 2005. Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking and Finance 29, 997–1015.

Simultaneity and Asymmetry of Returns and Volatilities in the Emerging Baltic State Stock Exchanges

Kurt Brännäs[a], Jan G De Gooijer[b], Carl Lönnbark[a] and Albina Soultanaeva[a]

[a] Department of Economics, Umeå University

[b] Department of Quantitative Economics, University of Amsterdam

Abstract

The paper suggests a nonlinear and multivariate time series model framework that enables the study of simultaneity in returns and in volatilities, as well as asymmetric effects arising from shocks. Using daily data 2000-2006 for the Baltic state stock exchanges and that of Moscow we find recursive structures with Riga directly depending in returns on Tallinn and Vilnius, and Tallinn on Vilnius. For volatilities both Riga and Vilnius depend on Tallinn. In addition, we find evidence of asymmetric effects of shocks arising in Moscow and in the Baltic state on both returns and volatilities.

Key Words: Time series, nonlinear, multivariate, finance, value at risk, portfolio allocation.

JEL Classification: C32, C51, G11, G12, G14, G15.

1 Introduction

This paper studies the joint evolution of returns and volatilities in the indices of the Baltic States stock exchanges, Riga (Latvia), Tallinn (Estonia), and Vilnius (Lithuania). The Baltic states have average annual growth rates in real GDP between 7.2 and 8.7 percent in 2000-2005, placing them among the fastest growing economies worldwide. These relatively small emerging marketplaces are geographically closely located. Besides sharing a common main owner, many of the largest traders are common to all three marketplaces. In fact, foreign institutional investors, predominately European ones, represent 40-47 percent of the market value in the Baltic stock markets. Foreign and domestic institutions together control about 90 percent of the market value. Given the common features of the markets their indices are likely to move together simultaneously.

The main motivation for this study is the importance of simultaneity for various investment and risk management decisions. Portfolio or fund managers, for example, often invest in several markets at the same time. This investment strategy may not provide the diversification and risk reduction that managers are seeking, if there are strong linkages between markets. In addition, risk managers need to understand the nature of cross market linkages in order to appropriately assess capital adequacy (Fleming et al., 1998).

Cross market linkages or information spillovers are of two types. The first is the common information that simultaneously affects expectations in more than one market. The second type of information spillovers is caused by cross-market hedging. Fleming et al. (1998) argue that information spillovers are strongest when linkages between markets are not limited by institutional constraints, and other practical considerations. Fazio (2007) argues that investors following an international diversification strategy may be exposed to unhedged risk when assuming that different countries are unrelated. He also finds that countries belonging to the same region are more likely to suffer from dependence in the case of extreme market movements. This implies that countries located in the same region may have stronger linkages than anticipated by investors.

A lesson from the current within-day trading literature concerning some other marketplaces is that information processing is very fast (e.g., Engle and Russell, 1998). Given the institutional setup of the Baltic state marketplaces it is likely that information transmission between these markets is virtually instantaneous. Even if there are unidirectional causations within the day, a study based on a daily sampling frequency cannot but find an average effect that may go both ways. The sampling frequency scenario is in fact a main motivation in macro-econometrics for employing structural systems which can incorporate simultaneous endogenous effects. Koch and Koch (1991) find simultaneity in returns within geographic regions but not across regions. More recently, Rigobon and Sack (2003) and others have reported on model-based studies allowing for simultaneity in returns.

Obviously, and perhaps more interestingly from a risk management point of view, there is also reason to expect simultaneous effects in volatilities. Rigobon and Sack (2003) were the first ones to find simultaneity in volatilities. But, as in the studies of De Wet (2006) and Lee (2006), the simultaneity arises in a very restrictive way, and only as a consequence of the simultaneity in returns. Gannon and Choi (1998) and Gannon (2004, 2005) detect simultaneity for some Asian markets using realized volatilities. Engle and Kroner (1995) suggested a related framework but focus theoretically on simultaneity in returns only.

Using three daily volume indices, Brännäs and Soultanaeva (2006) detected asymmetric effects in the series. Moreover, they demonstrated that good or bad news arriving from Russia (Moscow) have asymmetric impacts on the volatility transmissions for all indices under study. The model adopted was a univariate extension of an asymmetric ARMA (ARasMA) model introduced by Brännäs and De Gooijer (1994). Thus, each series was analyzed separately. Here, our main focus will be the joint modelling of, and the allowance for, simultaneity in both returns and volatilities along with asymmetry, and "Moscow" effects.

The model platform for the current study is the univariate ARasMA model combined with the asymmetric and quadratic GARCH of Brännäs and De Gooijer (1994, 2004). Brännäs and Soultanaeva (2006) extended this model class to allow for explanatory variables. The model is here to be given its first multivariate form and to allow for simultaneity in returns and volatilities separately. Notably, extensions of this type introduce additional parameters into an already richly parameterized model. Kroner and Ng (1998), De Goeij and Marquering (2005) and others discussed ways of parameterizing, in particular, the volatility functions for models to be estimable. To allow for simultaneity we will have to be restrictive in terms of correlation structure, lag lengths, and asymmetric effects.

The paper is organized as follows. In Section 2 we introduce the model and discuss some of its properties. In particular, we discuss the identifiability or uniqueness of estimation. Section 3 presents the estimator along with the employed stepwise model specification procedure. The section discusses testing against simultaneous, asymmetric, and Moscow effects. In addition, the use of the model for portfolio allocation and value at risk (VaR) studies are outlined. Section 4 presents the data-set. The empirical findings are given in Section 5. The final section concludes and relates our findings to other studies.

2 A Structural Vector ARasMA-asQGARCH Model

2.1 The Model

Consider an m-dimensional time series $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$. In this study $\{\mathbf{y}_t\}$ contains the variables of interest, i.e. the returns at time t of m stock market indices. The vector time series process $\{\mathbf{y}_t\}$ is assumed to be weakly stationary. Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{kt})'$ denote a vector of exogenous variables that may affect the process $\{\mathbf{y}_t\}$ like, within the context of this paper, the impact of news of the Russian stock exchange (RTS). To introduce the asymmetric structure of the proposed model we first need to define an m-dimensional vector discrete-time stochastic process generated by $\mathbf{u}_t = (u_{1t}, \ldots, u_{mt})'$ defined by

$$\mathbf{u}_t = \mathbf{H}_t^* \boldsymbol{\varepsilon}_t,$$

where $\{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \mathbf{I})$, $\mathbf{H}_t^* = \{h_{ij,t}^*\}$ $(i, j = 1, 2, \ldots, m)$, and $\mathcal{F}_{t-1}$ denotes the history of the time series up to and including time $t-1$. Hence, the conditional variance is $V(\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{H}_t^* \mathbf{H}_t^{*\prime} \equiv \mathbf{H}_t$. Then, asymmetries in the vector error process can be introduced as follows

$$\mathbf{u}_t^+ = \max(\mathbf{0}, \mathbf{u}_t) = \mathbf{H}_t^* \boldsymbol{\varepsilon}_t^+ \quad \text{and} \quad \mathbf{u}_t^- = \min(\mathbf{0}, \mathbf{u}_t) = \mathbf{H}_t^* \boldsymbol{\varepsilon}_t^-,$$

where $\boldsymbol{\varepsilon}_t^+ = \max(\mathbf{0}, \boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\varepsilon}_t^- = \min(\mathbf{0}, \boldsymbol{\varepsilon}_t)$. Now a simultaneous or structural vector ARasMA model can be defined as

$$\begin{aligned} \mathbf{A}_0 \mathbf{y}_t &= \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t + \sum_{i=1}^{q} \left(\mathbf{B}_i^+ \mathbf{u}_{t-i}^+ + \mathbf{B}_i^- \mathbf{u}_{t-i}^- \right) + \mathbf{c}_0 \\ &\quad + \sum_{i=0}^{r} \left(\mathbf{C}_i^+ \mathbf{x}_{t-i}^+ + \mathbf{C}_i^- \mathbf{x}_{t-i}^- \right), \end{aligned} \tag{1}$$

where $\mathbf{x}_t^+ = \max(\mathbf{0}, \mathbf{x}_t)$, and $\mathbf{x}_t^- = \min(\mathbf{0}, \mathbf{x}_t)$. Model (1) accounts for asymmetric effects unless for all i, $\mathbf{B}_i^+ = \mathbf{B}_i^-$ and $\mathbf{C}_i^+ = \mathbf{C}_i^-$. If appropriate, the threshold level for the process $\{\mathbf{x}_t\}$ may be set at another value than $\mathbf{0}$. Within the context of the present paper, the time series processes $\{\mathbf{x}_t^+\}$ and $\{\mathbf{x}_t^-\}$ represent positive and negative returns at time t in the RTS index. It is easy to see that the threshold levels in $\{\mathbf{u}_t^+\}$ and $\{\mathbf{u}_t^-\}$ can be accommodated by the vector of constants $\mathbf{c}_0$.

The $m \times m$ non-symmetric matrix $\mathbf{A}_0$ in (1) contains the simultaneity parameters,

$$\mathbf{A}_0 = \begin{pmatrix} 1 & a_{12}^0 & \cdots & a_{1m}^0 \\ a_{21}^0 & 1 & \cdots & a_{2m}^0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^0 & a_{m2}^0 & \cdots & 1 \end{pmatrix},$$

where an assumption of normalization has been imposed, i.e. coefficients along the diagonal are equal to 1. Assume $\mathbf{A}_0$ is nonsingular. Then the conditional mean (return) of $\{\mathbf{y}_t\}$ follows directly from the conditional reduced form of (1) as

$$\begin{aligned} E(\mathbf{y}_t|\mathcal{F}_{t-1}) &= \sum_{i=1}^{p} \mathbf{A}_0^{-1}\mathbf{A}_i\mathbf{y}_{t-i} + \sum_{i=1}^{q} \mathbf{A}_0^{-1}\left(\mathbf{B}_i^{+}\mathbf{u}_{t-i}^{+} + \mathbf{B}_i^{-}\mathbf{u}_{t-i}^{-}\right) \\ &\quad + \mathbf{A}_0^{-1}\mathbf{c}_0 + \sum_{i=0}^{r} \mathbf{A}_0^{-1}\left(\mathbf{C}_i^{+}\mathbf{x}_{t-i}^{+} + \mathbf{C}_i^{-}\mathbf{x}_{t-i}^{-}\right). \end{aligned}$$

Similarly, the conditional variance (volatility or risk) is given by

$$V(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{A}_0^{-1}\mathbf{H}_t(\mathbf{A}_0^{-1})'$$

from which, e.g., the conditional correlation matrix can be obtained. Various options are available to specify an asymmetric model for $\mathbf{H}_t$; see De Goeij and Marquering (2005). The specifications for $\mathbf{H}_t$ suggested by these authors contain off-diagonal elements. Thus there are conditional and possibly unconditional correlations among the elements of $\{\mathbf{u}_t\}$, and consequently among those of $\{\mathbf{y}_t\}$. There is no simultaneity in conditional volatility behavior in the sense that the conditional variance of, say, u_{it} would be a direct function of the corresponding conditional variance of u_{jt} $(i \neq j)$ in the same time period.

As we wish to have simultaneity in conditional volatility as an integral part of the model we need to consider an extension of the univariate asQGARCH model. One avenue that appears feasible is to view the structures of De Goeij and Marquering (2005) as "reduced forms". Note that structural forms may make economic sense but that only the reduced form gives the conditional variance interpretation. The situation resembles closely that of the simultaneous and reduced forms in classical macro-econometrics. Similarly, we view simultaneity to arise mainly due to the relatively low sampling frequency of one day while real trading occurs in continuous time, and partly due to identical actors on different stock exchanges.

Our general simultaneous specification for the conditional variance is very much in the same spirit as model (1). Given a vector time series process $\{\mathbf{z}_t\}$ of exogenous variables, the vector asQGARCH model for

$\mathbf{h}_t = vech(\mathbf{H}_t)$ is given by

$$\begin{aligned}\mathbf{D}_0\mathbf{h}_t &= \sum_{i=1}^{P}\mathbf{D}_i\mathbf{h}_{t-i} + \sum_{i=1}^{Q}\left(\mathbf{F}_i^+\mathbf{u}_{t-i}^+ + \mathbf{F}_i^-\mathbf{u}_{t-i}^-\right) + \sum_{i=1}^{Q}\mathbf{K}_i\mathbf{u}_{t-i}^{*,2} \\ &\quad + \mathbf{g}_0 + \sum_{i=0}^{R}\left(\mathbf{G}_i^+\mathbf{z}_{t-i}^+ + \mathbf{G}_i^-\mathbf{z}_{t-i}^-\right), \end{aligned} \tag{2}$$

where $\mathbf{g}_0$ is an $\frac{1}{2}m(m+1)\times 1$ vector of constants, $\mathbf{z}_t^+ = \max(\mathbf{0},\mathbf{z}_t)$, $\mathbf{z}_t^- = \min(\mathbf{0},\mathbf{z}_t)$, and the vector $\mathbf{u}_t^{*,2}$ has elements u_{it}^2 $(i = 1,\ldots,m)$. Within the context of the empirical analysis, the series $\{\mathbf{z}_t\}$ will enter (2) as the demeaned moving variance series of the RTS index; see Section 4 for more details on the construction of this series.

The reduced form of (2) is

$$\begin{aligned}\mathbf{h}_t &= \sum_{i=1}^{P}\mathbf{D}_0^{-1}\mathbf{D}_i\mathbf{h}_{t-i} + \sum_{i=1}^{Q}\mathbf{D}_0^{-1}\left(\mathbf{F}_i^+\mathbf{u}_{t-i}^+ + \mathbf{F}_i^-\mathbf{u}_{t-i}^-\right) + \sum_{i=1}^{Q}\mathbf{D}_0^{-1}\mathbf{K}_i\mathbf{u}_{t-i}^{*,2} \\ &\quad + \mathbf{D}_0^{-1}\mathbf{g}_0 + \sum_{i=0}^{R}\mathbf{D}_0^{-1}\left(\mathbf{G}_i^+\mathbf{z}_{t-i}^+ + \mathbf{G}_i^-\mathbf{z}_{t-i}^-\right)\end{aligned} \tag{3}$$

from which the corresponding $\mathbf{H}_t$ matrix can be obtained. The matrix $\mathbf{D}_0$ captures simultaneity, whereas the matrices $\mathbf{D}_i$ $(i \geq 1)$ are useful to represent persistence and possible cyclical features in the process $\{\mathbf{h}_t\}$. Also asymmetric effects are characterized through the matrices $\mathbf{F}_i^+$ $(\mathbf{F}_i^-)$ and $\mathbf{G}_i^+$ $(\mathbf{G}_i^-)$. Empirically, it is important to realize that the estimation of (3) may become infeasible with too generously parameterized specifications. Reducing lag lengths and introducing sparse matrix specifications are two ways of reducing the number of parameters; see Section 3 for a data-driven model specification procedure.

Various moment properties, and distributional results for univariate ARasMA models have been reported by Brännäs and De Gooijer (1994) and Brännäs and Ohlsson (1999), and for univariate ARasMA-quadratic GARCH models by Brännäs and De Gooijer (2004). Since $V(\mathbf{y}_t) = \mathbf{A}_0^{-1}E_{\mathcal{F}_{t-1}}(\mathbf{H}_t)(\mathbf{A}_0^{-1})' + V_{\mathcal{F}_{t-1}}[E(\mathbf{y}_t|\mathcal{F}_{t-1})]$, obtaining an explicit expression for the unconditional variance of $\{\mathbf{y}_t\}$ is a far from trivial problem.

2.2 Identification

We say that the system of simultaneous vector equations is identified when the parameters of the model can be uniquely estimated. Since estimation of the structural vector ARasMA-asQGARCH model will be in terms of its reduced form it is obvious that parameter matrices $\mathbf{A}_0$ and $\mathbf{D}_0$ play important roles. For instance, if $\mathbf{A}_0$ can be determined from lagged $\mathbf{y}_{t-i}$ parameters, all other parameters can be obtained uniquely. The situation is analogous for $\mathbf{D}_0$. The imposition of some sort of normalization restriction is necessary but not sufficient to achieve identification. A "traditional" solution is to impose long-run restrictions and/or sign restrictions on the parameters. However, within the context of our empirical analysis, we feel that these restrictions are difficult to defend. Instead we rely on a methodology proposed by Rigobon (2003) and Rigobon and Sack (2003) who showed that identification can be achieved if there is conditional heteroskedasticity in the data. The key idea is based on the movement of structural innovations $\{\mathbf{u}_t\}$ and the movement of the conditional covariances between them. The heteroskedasticity adds equations to the system, but also some unknowns. So, it is essential to impose some restrictions on the covariances to be able to use the variation in the second moments to solve the problem of identification. Rigobon (2003) derives necessary conditions for identification in case there are discrete regimes in the variances of the structural shocks. In our structural vector model, the variances of the shocks are allowed to evolve in a continuous manner. Thus giving rise to a continuum of regimes for identifying the system.

3 Estimation and Model Use

Given a multivariate normality assumption on $\{\boldsymbol{\varepsilon}_t\}$ the prediction error

$$\mathbf{y}_t - E(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{A}_0^{-1}\mathbf{u}_t = \mathbf{A}_0^{-1}\mathbf{H}_t^*\boldsymbol{\varepsilon}_t \equiv \mathbf{v}_t$$

is conditionally $N(\mathbf{0}, \boldsymbol{\Gamma}_t)$ distributed with $\boldsymbol{\Gamma}_t = \mathbf{A}_0^{-1}\mathbf{H}_t(\mathbf{A}_0^{-1})'$; recall (3). Here, $\mathbf{H}_t$ is the conditional variance expression in reduced form, containing among other things the $\mathbf{D}_0$ matrix. Given observations up till time T, the log-likelihood function takes the form

$$\begin{aligned}\ln L &\propto -\frac{1}{2}\sum_{t=s}^{T}\ln|\mathbf{\Gamma}_t| - \frac{1}{2}\sum_{t=s}^{T}\mathbf{v}_t'\mathbf{\Gamma}_t^{-1}\mathbf{v}_t \\ &\propto (T-s)\ln|\mathbf{A}_0| - \frac{1}{2}\sum_{t=s}^{T}\left(\ln|\mathbf{H}_t| + \mathbf{u}_t'\mathbf{H}_t^{-1}\mathbf{u}_t\right),\end{aligned}$$

where $s = \max(p, q, r) + 1$. For practical quasi maximum likelihood estimation we use the RATS 6.0 package and employ robust standard errors.

To obtain the final model specification we advocate the following stepwise procedure.

1. Univariate ARasMA-asQGARCH models containing specifications for both mean returns and conditional variance are first estimated. Select models that minimize AIC or some other appropriate model selection criterion. Thus, we implicitly assume that there are no interactions between the series.

2. Using results from step 1 introduce simultaneity in the structural form, i.e. add $\mathbf{A}_0$. Consider thereafter the expansion to non-diagonal matrices in the returns expression. Choose the specification that minimizes AIC. The $\mathbf{A}_0$ is the final parameter matrix to be reduced. For this step the volatility functions obtained in step 1 are taken as given, but $\{\hat{\mathbf{u}}_t\}$ changes in the iterative steps.

3. Using results from steps 1 and 2 introduce simultaneity in the volatility function, i.e. add $\mathbf{D}_0$. Consider thereafter the expansion to non-diagonal matrices in the volatility expression. Choose the specification that minimizes AIC. The $\mathbf{D}_0$ is the final parameter matrix to be reduced and the $\{\hat{\mathbf{u}}_t\}$-sequence are taken as given from step 2.

4. In a final step all parameters are estimated jointly.

Given the estimated model, it is of interest to test hypotheses about simultaneity, asymmetry, and the Moscow effect. Given the likelihood framework and our specification procedure, Wald and likelihood ratio (LR) test statistics are relatively easy to implement.

We first consider tests of simultaneity and do so in terms of the $\mathbf{A}_0$ matrix. The reasoning with respect to $\mathbf{D}_0$ is analogous. We say that there is a simultaneous effect between markets i and j if $(\mathbf{A}_0)_{ij} \neq 0$ and $(\mathbf{A}_0)_{ji} \neq 0$. When $(\mathbf{A}_0)_{ij} \neq 0$ but $(\mathbf{A}_0)_{ji} = 0$ there is a recursive structure and causation is unidirectional from market j to market i. When $(\mathbf{A}_0)_{ij} = (\mathbf{A}_0)_{ji} = 0$ there is no causation between returns. When all off-diagonal elements equal zero $\mathbf{A}_0 = \mathbf{I}$ and the structural and reduced forms are identical.

Next we consider testing against asymmetric effects and do so in terms of the $\mathbf{B}_i^+$ and $\mathbf{B}_i^-$ matrices. We may form $\mathbf{B}_i^\nabla = \mathbf{B}_i^+ - \mathbf{B}_i^-$ $(i = 1, \ldots, q)$, and test whether this matrix is equal to zero or whether it is nonzero. We then make no distinction between the case of both matrices having nonzero parameters $(\mathbf{B}_i^+)_{ij}$ and $(\mathbf{B}_i^-)_{ij}$ in all places and the case where, say, $(\mathbf{B}_i^-)_{ij} = 0$. Testing against asymmetric effects of Moscow is in terms of the parameter matrices $\mathbf{C}_i^+$ and $\mathbf{C}_i^-$ $(i = 1, \ldots, r)$. For asymmetric effects in volatility the parameter matrices $\mathbf{F}_i^+$ and $\mathbf{F}_i^-$ as well as $\mathbf{G}_i^+$ and $\mathbf{G}_i^-$ are focused.

For no effects of Moscow on returns all matrices $\mathbf{C}_i^+$ and $\mathbf{C}_i^-$ must be identical to a zero matrix, while for volatility all $\mathbf{G}_i^+$ and $\mathbf{G}_i^-$ must be zero.

When we wish to use or, as here, evaluate the model in financially interesting and meaningful ways, portfolio allocation and VaR measures are of obvious interest. Two problems both stemming from the use of index series arise; how to get back to the index and what price related to the index should we consider.

First, the index is determined from the inverse of the change variable $y_{it} = 100 \ln(I_{it}/I_{it-1})$, i.e. as $I_{it} = I_{it-1} \exp(y_{it}/100)$ for stock market i. We get $E(I_{it}|\mathcal{F}_{t-1}) = I_{it-1} E(\exp(y_{it}/100)|\mathcal{F}_{t-1}) \approx I_{it-1}(1 + E(y_{it}|\mathcal{F}_{t-1})/100)$ where the first order approximation of the exponential function is reasonable for the small values of $y_{it}/100$. Using the same first order approximation we get $V(\mathbf{I}_t|\mathcal{F}_{t-1}) = \mathbf{I}^{\circ}_{t-1}V(\mathbf{y}_t|\mathcal{F}_{t-1})\mathbf{I}^{\circ}_{t-1}/100^2$, where $\mathbf{I}^{\circ}_t$ is a matrix with elements I_{it} on the diagonal and zeroes elsewhere. These expressions are useful if we wish to forecast the index and to give its forecast variance. Second, trading is not directly in terms of the index. The presence of index funds and standard options tied to the index are reasonable justifications for using the index as a price. The chosen approach is to use the return series as is and then emphasize the return as an indicator of market risk (e.g., McNeil and Frey, 2000).

For portfolio allocation we adopt the tangency portfolio (e.g., Campbell et al., 1997, ch 5). At time $T+1$ we have

$$\mathbf{a}_{T+1} = V^{-1}(\mathbf{y}_{T+1}|\mathcal{F}_T) \cdot \left[E(\mathbf{y}_{T+1}|\mathcal{F}_T) - R_f\mathbf{1}\right]/A,$$

where $A = \mathbf{1}'V^{-1}(\mathbf{y}_{T+1}|\mathcal{F}_T) \cdot [E(\mathbf{y}_{T+1}|\mathcal{F}_T) - R_f\mathbf{1}]$, R_f is the risk free rate, and $\mathbf{1}$ is a column vector of ones. Hence, $\mathbf{1}'\mathbf{a}_{T+1} = 1$. For the VaR-measure under normality, a time invariant allocation vector $\mathbf{a}$, and a probability α, Gourieroux and Jasiak (2001, ch 16) give:

$$R_{T+1} = -\mathbf{a}'E(\mathbf{y}_{T+1}|\mathcal{F}_T) + \Phi^{-1}(1-\alpha)\left[\mathbf{a}'V(\mathbf{y}_{T+1}|\mathcal{F}_T)\mathbf{a}\right]^{1/2}.$$

This VaR measure is in terms of returns; one in terms of indices can also be devised by simply replacing $\mathbf{y}_{T+1}$ by $\mathbf{I}_{T+1}$ and using the expressions given above. Using shock scenarios in terms of the $\mathbf{u}_t$ vector or in terms of $\mathbf{x}_t^{+/-}$ and $\mathbf{z}_t^{+/-}$, the $\mathbf{a}_{T+1}$ and R_{T+1} can be calculated and then evaluated and subjected to comparisons. To cast light on effects of simultaneity, the univariate models can be compared to the simultaneous model system in terms of the portfolio or VaR metrics either as above or over some historical period. Note, that both measures are subject to sampling variation in estimated mean return and risk functions. Britten-Jones (1999) and others have discussed the variation in allocation weights, while Christoffersen and Gonçalves (2005) among others have discussed the issue for VaR measures.

4 Data

The data used in this paper are capitalization weighted daily stock price indices of the Estonian (Tallinn, TALSE), Latvian (Riga, RIGSE), Lithuanian (Vilnius, VILSE) and Russian (Moscow, RTS) stock markets. All prices are transformed into Euros from local currencies, except for Estonia where stock market trading is in Euro. The data-set covers January 3, 2000 to August 16, 2006, for a total of $T = 1729$ observations, cf. Figure 1 for the three Baltic indices. Both indices and exchange rates are collected from DataStream. The irregularity in the summer of 2001 in the Riga index (RIGSE) is due to a power struggle in its largest company (Latvijas Gaze). Instead of elaborating on modelling to contain this irregular period, the Riga series is adjusted in the following simplistic way: For a speculation period from July 25 to September 3, 2001, observations are replaced by interpolated values. The returns of Moscow serve as the x_t variables in (1). For the z_t of the conditional variance function in (3) we construct a new series by obtaining moving variances for a window length of 10 observations. The sample mean is 4.65 with a variance of 28.83. The z_t series that enter the conditional variance function are demeaned moving variance series; the threshold is then set at zero. The z^+ then takes on positive values and is indicative of high-risk, and z^- in a corresponding way takes on negative values and indicates a lower risk in Moscow.

Due to some differences in holidays for the involved countries the series have different shares of days for which index stock price are not observable. Linear interpolation was used to fill the gaps for all series. The resulting series are then throughout for a common trading week. All returns are calculated as $y_t = 100 \cdot \ln(I_t/I_{t-1})$, where I_t is the daily price index. Table 1 reports descriptive statistics for the daily returns. The Ljung-Box statistics for 10 lags (LB_{10}) indicate significant serial correlations. The large kurtoses for Riga, Tallinn and Vilnius indicate leptokurtic densities. Table 2 presents cross correlations for the Baltic return series and for a squared returns. Table 3 gives auto and lagged cross correlations. For instance, the table indicates that Tallinn is positively affected by Vilnius both within the day and with up to three lags. There appears to be no impact from Riga.

Figure 2 gives scatterplots for pairs of returns series with a nonparametric regression line (LOWESS default settings in RATS 6.0). Visual

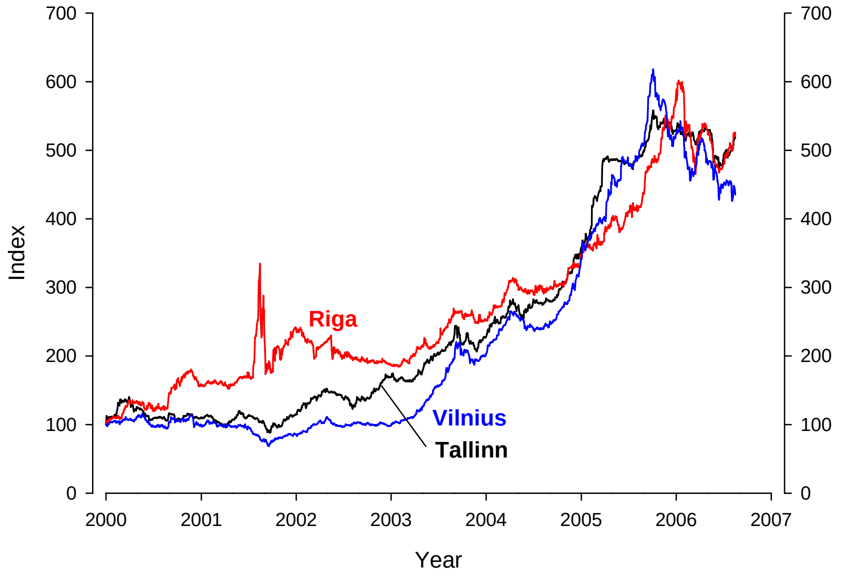

Figure 1: Indices of the Baltic stock exchanges (December 31, 1999 = 100).

Table 1: Descriptive statistics for return series.

Exchange	Mean	Variance	Min/Max	Skewness	Kurtosis	LB_{10}
Riga	0.10	1.77	-9.27/10.29	0.18	10.72	45.93
Tallinn	0.10	1.05	-5.87/12.02	1.09	15.94	51.43
Vilnius	0.09	1.05	-12.12/5.32	-0.91	13.82	46.87
Moscow	0.12	4.93	-11.92/10.23	-0.47	3.27	16.37

Note: LB_{10} is the Ljung-Box statistic evaluated at 10 lags.

Table 2: Cross correlations for Baltic stock markets returns and squared returns.

	Returns			Squared Returns		
	Riga	Tallinn	Vilnius	Riga	Tallinn	Vilnius
Riga	1			1		
Tallinn	0.134	1		0.161	1	
Vilnius	0.141	0.208	1	0.023	0.032	1

Table 3: Auto and cross correlations for Baltic stock markets returns (in the order Riga, Tallinn and Vilnius). Significant entries are indicated by signs and subindex indicates lag.

$$\begin{pmatrix} 1 & + & + \\ . & 1 & + \\ . & + & 1 \end{pmatrix}_0, \begin{pmatrix} - & . & . \\ . & . & + \\ . & + & + \end{pmatrix}_1, \begin{pmatrix} . & . & + \\ . & . & + \\ . & . & + \end{pmatrix}_2, \begin{pmatrix} + & . & . \\ . & . & + \\ . & . & . \end{pmatrix}_3, \begin{pmatrix} - & . & . \\ . & . & . \\ . & . & + \end{pmatrix}_4$$

inspection indicates that there is weak dependence between Riga and Tallinn for the majority of observations, while for the other plots there appear to be positive relationships.

5 Results

The empirical results are presented first in terms of the return function and later in terms of the volatility function. Table A contains estimated univariate models. The empirical specifications are obtained by the steps outlined in Section 3.

For the return function of $\{\mathbf{y}_t\}$, cf. eq (1), when returns are in the

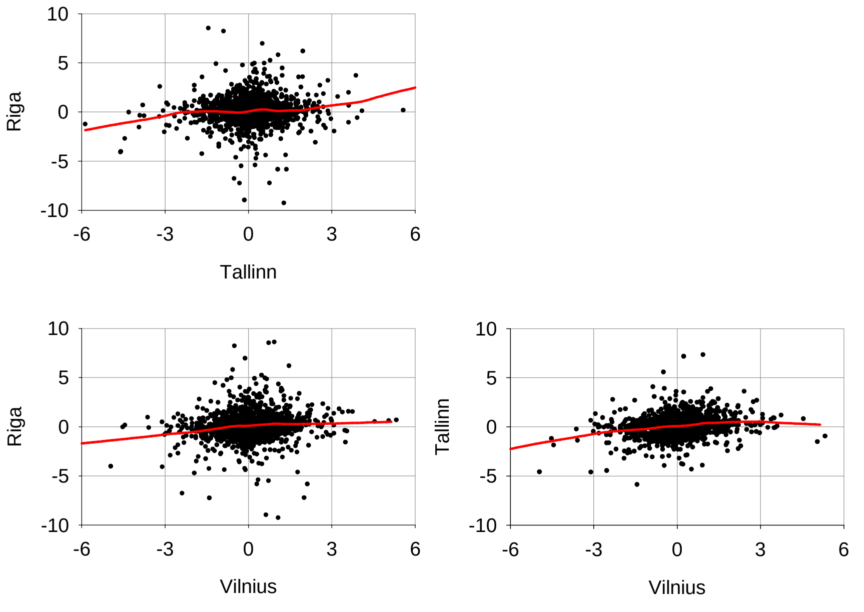

Figure 2: Cross plots for Baltic returns series. One negative outlier for Vilnius is outside the figure and three positive ones for Tallinn.

order Riga, Tallinn and Vilnius, the estimated function is

$$
\begin{aligned}
&\begin{pmatrix} 1 & \underset{(0.034)}{-0.06} & \underset{(0.043)}{-0.09} \\ 0 & 1 & \underset{(0.025)}{-0.11} \\ 0 & 0 & 1 \end{pmatrix} \hat{\mathbf{y}}_t = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \underset{(0.023)}{0.06} \end{pmatrix} \hat{\mathbf{y}}_{t-2} \\
&+ \begin{pmatrix} \underset{(0.050)}{-0.17} & 0 & 0 \\ 0 & \underset{(0.044)}{0.24} & 0 \\ 0 & 0 & \underset{(0.048)}{0.15} \end{pmatrix} \hat{\mathbf{u}}_{t-1}^{+} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \underset{(0.039)}{0.12} & 0 \\ 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{u}}_{t-2}^{+} \\
&+ \begin{pmatrix} 0 & 0 & 0 \\ \underset{(0.023)}{0.07} & \underset{(0.048)}{0.09} & \underset{(0.026)}{0.07} \\ 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{u}}_{t-1}^{-} + \begin{pmatrix} 0 & \underset{(0.046)}{0.12} & \underset{(0.041)}{0.08} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{u}}_{t-2}^{-} \\
&+ \begin{pmatrix} \underset{(0.043)}{0.21} \\ \underset{(0.026)}{0.12} \\ \underset{(0.027)}{0.14} \end{pmatrix} + \begin{pmatrix} \underset{(0.021)}{0.06} \\ 0 \\ 0 \end{pmatrix} x_t^{+} + \begin{pmatrix} 0 \\ 0 \\ \underset{(0.018)}{0.04} \end{pmatrix} x_{t-1}^{+} \\
&+ \begin{pmatrix} \underset{(0.024)}{0.07} \\ \underset{(0.011)}{0.09} \\ \underset{(0.016)}{0.12} \end{pmatrix} x_t^{-} + \begin{pmatrix} 0 \\ \underset{(0.012)}{0.03} \\ 0 \end{pmatrix} x_{t-1}^{-} + \begin{pmatrix} 0 \\ \underset{(0.014)}{0.02} \\ 0 \end{pmatrix} x_{t-2}^{-}.
\end{aligned}
$$

With respect to simultaneity, the $\hat{\mathbf{A}}_0$ matrix indicates a recursive structure; the returns of the Riga index depends within the day positively on both the index returns of Tallinn and Vilnius, while returns in Tallinn are positively influenced by those of Vilnius. Riga returns have no impact on the returns of neither Tallinn nor Vilnius, and Tallinn returns have no influence on those of Vilnius. The only lagged influence arises for Vilnius at lag two, cf. the $\hat{\mathbf{A}}_2$ matrix. For Riga returns Moscow has a quite symmetric and positive effect within the day. For Tallinn we instead find asymmetric and negative effects spread over lags $0-2$, and for Vilnius negative shocks out of Moscow appear to have larger impact than positive shocks. For shocks arising in the three Baltic stock exchanges we find that a positive shock in Riga at lag one has a negative impact on current returns, and in addition negative lag two shocks of Tallinn and Vilnius have negative effects. Positive shocks in Tallinn have stronger effects than equally sized negative shocks, and there are negative shocks of both Riga and Vilnius at lag 2. The off-diagonal elements of lagged shocks suggests that there are some shock-spillovers; Riga returns are negatively influenced by Tallinn and Vilnius shocks at lag two, while Tallinn is impacted by Riga and Vilnius shocks at lag one.

The estimated volatility function has the form

$$
\begin{pmatrix} 1 & \underset{(0.004)}{-0.01} & 0 \\ 0 & 1 & 0 \\ 0 & \underset{(0.016)}{0.03} & 1 \end{pmatrix} \hat{\mathbf{h}}_t = \begin{pmatrix} \underset{(0.006)}{0.95} & 0 & 0 \\ 0 & \underset{(0.009)}{0.93} & 0 \\ 0 & 0 & \underset{(0.029)}{0.82} \end{pmatrix} \hat{\mathbf{h}}_{t-1}
$$

$$
+ \begin{pmatrix} \underset{(0.018)}{-0.02} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \underset{(0.025)}{-0.12} & \underset{(0.035)}{0.27} \end{pmatrix} \hat{\mathbf{u}}^{+}_{t-1} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \underset{(0.016)}{0.15} & 0 \\ 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{u}}^{+}_{t-2}
$$

$$
+ \begin{pmatrix} \underset{(0.076)}{0.37} & 0 & 0 \\ 0 & \underset{(0.017)}{-0.15} & 0 \\ 0 & 0 & \underset{(0.033)}{-0.26} \end{pmatrix} \hat{\mathbf{u}}^{-}_{t-1} + \begin{pmatrix} \underset{(0.073)}{-0.29} & 0 & \underset{(0.008)}{-0.06} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{u}}^{-}_{t-2}
$$

$$
+ \begin{pmatrix} \underset{(0.039)}{0.36} & 0 & 0 \\ 0 & \underset{(0.037)}{0.13} & 0 \\ 0 & \underset{(0.007)}{0.03} & \underset{(0.033)}{0.12} \end{pmatrix} \hat{\mathbf{u}}^{*,2}_{t-1} + \begin{pmatrix} \underset{(0.037)}{-0.31} & 0 & 0 \\ 0 & \underset{(0.034)}{-0.14} & \underset{(0.001)}{-0.01} \\ 0 & 0 & \underset{(0.028)}{-0.13} \end{pmatrix} \hat{\mathbf{u}}^{*,2}_{t-2}
$$

$$
+ \begin{pmatrix} \underset{(0.011)}{0.02} \\ \underset{(0.004)}{-0.02} \\ \underset{(0.026)}{0.07} \end{pmatrix} + \begin{pmatrix} \underset{(0.000)}{-0.002} \\ 0 \\ \underset{(0.006)}{0.04} \end{pmatrix} z_t^{+} + \begin{pmatrix} \underset{(0.017)}{0.08} \\ \underset{(0.012)}{0.04} \\ 0 \end{pmatrix} z_t^{-}
$$

$$+\begin{pmatrix} 0 \\ 0 \\ \underset{(0.006)}{-0.04} \end{pmatrix} z_{t-1}^{+} + \begin{pmatrix} \underset{(0.016)}{-0.08} \\ \underset{(0.012)}{-0.04} \\ \underset{(0.004)}{0.01} \end{pmatrix} z_{t-1}^{-}.$$

Only two elements in $\hat{\mathbf{D}}_0$ are significant, the volatility of Vilnius depends negatively but weakly on that of Tallinn in the same time period, while Riga depends positively on Tallinn. As expected volatilities are quite persistent, cf. the $\hat{\mathbf{D}}_1$-matrix estimates. In the very short term (within the day) a higher than average Moscow risk marginally reduces risk in Riga, while the effect is an enhancing one for Vilnius. Already after one day there appears to remain little impact of Moscow risk for Vilnius. This is also true for negative shocks in all three stock markets.

The conditional covariances are very small and insignificantly estimated as $\mathbf{H}_{t,1,2} = 0.003$ (s.e. $= 0.023$), $\mathbf{H}_{t,1,3} = 0.000$ (0.033) and $\mathbf{H}_{t,2,3} = 0.000$ (0.025).

The model evaluation phase considers formal tests against simultaneity in returns and in risk as well as tests against asymmetric effects arising from Moscow or from the innovations of the model system. As a first but informal test supporting the joint models rests on the likelihoods under the univariate models and the joint model; the likelihood ratio statistic is then $LR = 181.8$. Table 4 summarizes the formal test results and also gives the serial correlation properties and the goodness-of-fit for the model. The Wald tests are all significant with p-values less than 0.02. There is then evidence of simultaneity as well as of asymmetric effects. When it comes to serial correlation properties in standardized and squared standardized residuals there appears to be remaining serial correlation in only one series, the standardized residuals of Vilnius. The standardized residuals are nonnormal and leptokurtic.

Next, we consider the estimated volatility functions in some more detail in Figures 3-4. Figure 3 shows the estimated $\mathbf{H}_{t,i,i}$ functions for the final part of the series. It is quite clear from this figure that the volatilities of Riga and Vilnius are larger than those of Tallinn. This pattern reenforces the sample variance ordering of Table 1. The esti-

Table 4: Simultaneity and asymmetry tests together with model evaluation measures.

Hypothesis	Wald	df	Measure	Riga	Tallinn	Vilnius
Simultaneity-Returns	27.0	3	LB_{10}	10.08	5.82	22.75
Simultaneity-Risk	7.81	2	LB^2_{10}	11.77	1.63	1.14
Asymmetry-Return-Moscow	160.9	6	Skewness	0.47	0.54	-0.30
Asymmetry-Return-Innovation	74.4	8	Kurtosis	4.33	6.31	6.06
Asymmetry-Risk-Moscow	92.8	6	JB	1403.7	2936.8	2659.2
Asymmetry-Risk-Innovation	6033	7	R^2	0.05	0.18	0.06

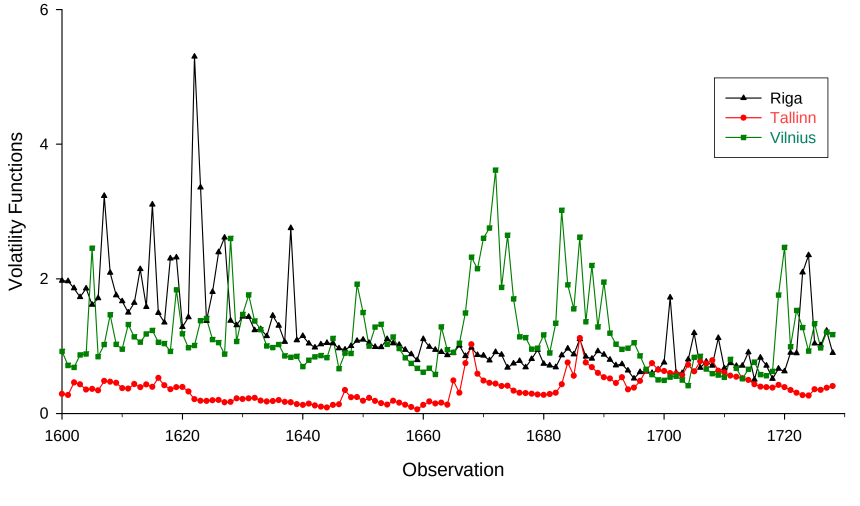

Figure 3: Estimated volatility functions for the final part of the sample period.

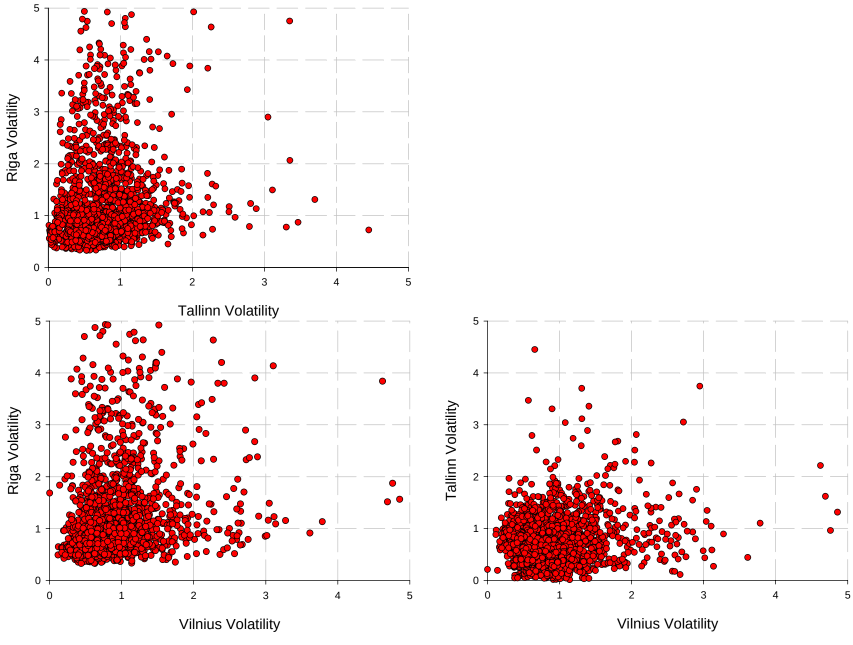

Figure 4: Plots of estimated volatilities (some outlying volatilities fall outside the graphs).

mated volatility functions are positively correlated, cf. Figure 4. Since covariance estimates $\mathbf{H}_{t,i,j}$ between the innovations of stock exchanges are very small the resulting time-varying conditional correlations are also very small and always smaller than 0.05. The implied estimated conditional correlations between $\{\mathbf{y}_t\}$ variables are much larger and also positive throughout, cf. Figure 5. Average conditional correlations are relatively close to the sample correlations of Table 2.

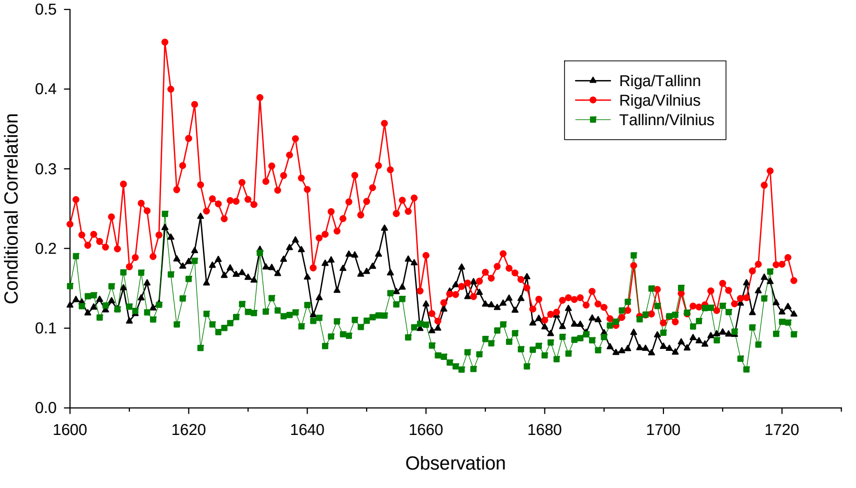

Figure 5: Estimated conditional correlations between the returns of the stock markets for the final part of the sample period.

Table 5: Portfolio and VaR effects of shocks in innovations and Moscow (Joint), together with a univariate model (Single) case. The VaR is based on probability 0.025 and a portfolio with weights 0.333 for each index (VaR-A) and with the weights obtained in the Base case (VaR-B).

	Portfolio Allocation						VaR			
	Joint			Single			A		B	
	Riga	Tallinn	Vilnius	Riga	Tallinn	Vilnius	Joint	Single	Joint	Single
Base case	0.24	0.66	0.10	0.32	0.50	0.18	1.23	0.91	1.66	0.83
Shock-Riga	0.27	0.64	0.09	0.19	0.60	0.21	1.19	1.15	1.65	0.98
-Tallinn	0.30	0.54	0.16	0.35	0.45	0.19	1.23	0.99	1.64	1.06
-Vilnius	0.26	0.72	0.02	0.37	0.58	0.05	1.42	0.99	2.02	0.81
-Moscow (x)	0.23	0.67	0.10	0.31	0.51	0.18	1.25	0.90	1.67	0.82
-Moscow (z)	0.24	0.62	0.14	0.27	0.50	0.23	1.36	1.07	1.77	1.03

Portfolio allocations and VaR measures one-step-ahead are depicted in Table 5. These measures are based on forecast equations

$$
\begin{aligned}
E(\mathbf{y}_{T+1}|\mathcal{F}_T) &= \hat{\mathbf{A}}_0^{-1}\left[\hat{\mathbf{A}}_2\mathbf{y}_{T-1} + \sum_{i=1}^{2}\left(\hat{\mathbf{B}}_i^{+}\hat{\mathbf{u}}_{T+1-i}^{+} + \hat{\mathbf{B}}_i^{-}\hat{\mathbf{u}}_{T+1-i}^{-}\right)\right. \\
&\qquad \left. +\hat{\mathbf{c}}_0 + \sum_{i=0}^{2}\left(\hat{\mathbf{C}}_i^{+}\mathbf{x}_{T-i}^{+} + \hat{\mathbf{C}}_i^{-}\mathbf{x}_{T-i}^{-}\right)\right] \\
V(\mathbf{y}_{T+1}|\mathcal{F}_T) &= \hat{\mathbf{A}}_0^{-1}\hat{\mathbf{H}}_{T+1}(\hat{\mathbf{A}}_0^{-1})'
\end{aligned}
$$

and depend on the histories of $\mathbf{y}_t$, $\hat{\mathbf{u}}_t$, and $\mathbf{x}_t$ for the conditional return and additionally on the histories of $\mathbf{H}_t$ and $\mathbf{z}_t$ for the conditional volatility. Since the impact of Moscow is in the same period we set future values (x_{T+1} and z_{T+1}) for Moscow close to their values at the end of the series, i.e. as $x_{T+1}^{+} = 0.1$ and $z_{T+1}^{-} = -4$. This is the Base case design. For the portfolio allocation exercise the risk free rate is set at 1.07, which is the level of the Euro market government bond yield by the end of the sample period.

The allocation for the Tallinn stock exchange is 0.66, while 0.24 of the portfolio should be placed in Riga and 0.10 in Vilnius. Using the same setup but using instead the univariate models (Single) of Table A, gives a much lower allocation for Tallinn and higher ones for both Riga and Vilnius.[1] The two model forms differ in simultaneity but also with respect to other features of the dynamic model. Therefore, we cannot infer with certainty that the differences are due solely to simultaneous effects. The VaR measures for probability 0.025 are for the simultaneous model with equal weights 1.23 and for the univariate models 0.91. For the weights obtained with the weights of the Base case we get 1.66 and 0.82, respectively.

To study the sensitivity of the Base case results we next shock the individual elements of $\hat{\mathbf{u}}_T$ (the final residuals are individually multiplied

[1] In shocking the stock markets, note that the residuals of the joint and univariate models differ both in sizes and signs. The underlying sizes of residuals in the univariate models have not been changed but shocks are throughout in the direction of the joint model.

by a factor 3). For shocks in the Tallinn and Vilnius stock markets the allocations for these markets are reduced. Figure 6 illustrates this for an increasingly negative shock in Tallinn. With a decrease in the Tallinn weight comes relatively more weight for Riga than for Vilnius. The allocations obtained using the univariate models differ from those based on the joint model, mainly such that the weights for Riga and Vilnius are larger and those for Tallinn are smaller.

We also consider shocks arising in Moscow returns (x^+_{T+1} is set to 1). This appears to have only minor impact. For Moscow risk we change from $z^-_{T+1} = -4$ to $z^+_{T+1} = 4$ and note an increase for Vilnius and a reduction for Tallinn allocations.

The VaR measure changes little for shocks in Tallinn but responds more to shocks in Vilnius and in Moscow risk. The VaR:s based on the univariate models are smaller than the corresponding measures for the joint model. When the weights of the Base case are used the VaR:s increase markedly throughout. Figure 6 studies the impacts on VaR of Moscow shocks in more detail. Changes in risk have rather small effects, while Moscow return changes have a more sizeable and asymmetric effect.

6 Conclusion

The paper has introduced simultaneity into a multivariate and nonlinear time series model framework to study jointly the indices of the Baltic states stock exchanges. Unlike previous studies (e.g., Rigobon and Sack, 2003, De Wet, 2006, Lee, 2006), we allow for simultaneity in returns and volatility separately. The model allows us to capture "within a day" information transmission between the stock markets under study. Since information transmission between markets is virtually instantaneous (e.g., Engle and Russell, 1998) a study based on daily sampling frequency should take into account simultaneous reactions to movements in other relevant assets or markets. Moreover, the model is able to capture asymmetric impacts of lagged positive and negative shocks on returns and volatility processes. We argue that measuring

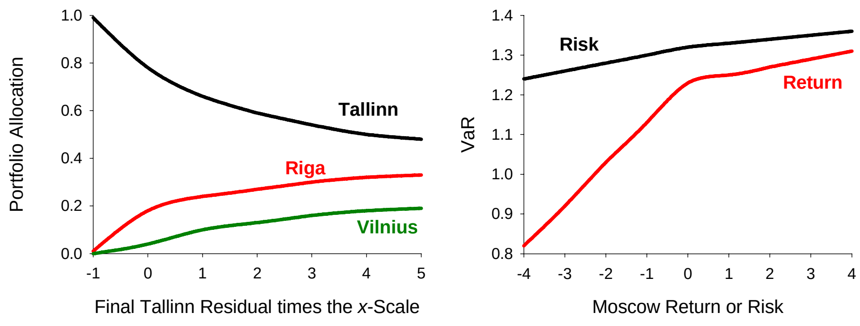

Figure 6: Allocations after shocking the final negative residual for Tallinn (left exhibit, a value on the x-scale larger than +1 means a larger negative shock). VaR effects of shocks to Moscow return and risk (right exhibit).

simultaneous and asymmetric spillovers is important for a number of reasons, including optimal portfolio allocation and risk management.

In summary, the empirical analysis provides support for the simultaneity in return and volatility. Accounting for simultaneity is of particular importance for markets located in the same geographic region or closely related due to institutional structure or other practical considerations as for example common trading platform. Given the fact that investors diversify their holding across markets in order to reduce the risk of the portfolio, accounting for information which simultaneously alters the expectations of different markets is important for asset allocation and risk management strategies.

Empirically, we illustrate the importance of simultaneity with respect to Baltic stock markets. In these closely related markets simultaneity is likely to arise due to geographic proximity, common institutional setup as well as common large traders, among other things. We found strong evidence of simultaneous effects in both returns and volatility. In returns, Riga is dependent on the indices of Tallinn and Vilnius, Tallinn is dependent on Vilnius, while Vilnius is not influenced by the other two markets. For volatility, we find within a day spillovers from Tallinn to both Riga and Vilnius. In addition, we found asymmetric effects of Moscow returns on the index returns in the Baltic exchanges, and asymmetric effects of Moscow risk on volatilities.

To illustrate the importance of simultaneous interaction between markets we obtain the portfolio allocations and value at risk measures for the multivariate and univariate models. Portfolio allocation results indicate that optimal portfolio weights are more sensitive to shocks when simultaneity is not accounted for. VaR measures indicate that the variability in losses that may occur due to shocks to the market is larger when simultaneity is not accounted for.

The simultaneous and dynamic econometric model generalizes previous univariate models by allowing for simultaneity but also for cross-effects of innovations. As in any simultaneous model we can therefore talk about direct, indirect and total effects in the return and volatility functions. The direct effects can be seen in the estimation results, while the portfolio and value at risk results build on total effects. To estimate the model we employ full information maximum likelihood. The suggested stepwise specification procedure resulted in a model with important deviations from corresponding univariate models. Estimation of the final model does not result in numerical problems despite the fact that the model is quite richly parametrized.

Acknowledgements

The financial support from the Wallander-Hedelius foundation to the first three authors and from the Nordea foundation to Albina Soultanaeva are gratefully acknowledged. A previous version of this paper was presented at the Nordic Econometric Meeting and EC^2, 2007.

Appendix

Table A: Estimation results for univariate models.

	Riga				Tallinn				Vilnius			
Variables	Return		Risk		Return		Risk		Return		Risk	
y_{t-2}									0.057	0.021		
u^+_{t-1}	-0.146	0.048	-0.072	0.018	0.252	0.041			0.162	0.045	0.273	0.030
u^+_{t-2}					0.117	0.037	0.021	0.022				
u^+_{t-3}												
u^-_{t-1}			0.394	0.071	0.119	0.046	-0.190	0.181			-0.279	0.023
u^-_{t-2}			-0.283	0.064								
h_{t-1}			0.944	0.005			0.917	0.009			0.829	0.024
u^2_{t-1}			0.389	0.034			0.113	0.034			0.093	0.031
u^2_{t-2}			-0.322	0.031			-0.135	0.031			-0.113	0.026
x, z^+_t	0.050	0.021	-0.001	0.001							0.034	0.005
x, z^+_{t-1}									0.046	0.0167	-0.032	0.005
x, z^-_t	0.105	0.021	0.121	0.0167	0.120	0.011	0.046	0.012	0.126	0.015	0.007	0.004
x, z^-_{t-1}			-0.114	0.0167	0.046	0.012	-0.050	0.012				
x, z^-_{t-2}					0.029	0.013						
Constant	0.177	0.033	0.079	0.012	0.114	0.027	-0.035	0.004	0.141	0.027	0.004	0.020
AIC	2086.9				1164.5				1446.8			
$\ln L, R^2$	-1029.5		0.03		-566.87			0.16	-709.41			0.06
LB_{10}	10.84		8.83		7.01		1.53		21.57		1.53	
Skew, Kurt, JB	0.43	5.60	2303.5		0.439	6.86	3446.6		-0.23	6.48	3030.7	

References

Brännäs K, De Gooijer JG. 1994. Autoregressive-asymmetric moving average models for business cycle data. *Journal of Forecasting* **13**, 529-544.

Brännäs K, De Gooijer JG. 2004. Asymmetries in conditional mean and variance: modelling stock returns by asMA-asQGARCH. *Journal of Forecasting* **23**, 155-171.

Brännäs K, Ohlsson H. 1999. Asymmetric time series and temporal aggregation. *Review of Economics and Statistics* **81**, 341-344.

Brännäs K, Soultanaeva A. 2006. Influence of news in Moscow and New York on returns and risks on Baltic state stock indices. Umeå Economic Studies 696, Umeå University.

Britten-Jones M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* **LIV**, 655-671.

Campbell JY, Lo AW, MacKinlay AC. 1997. *The Econometrics of Financial Markets.* 2nd edition, Princeton University Press, Princeton.

Christoffersen P, Gonçalves S. 2005. Estimation in financial risk management. *Journal of Risk* **7**, 1-28.

De Goeij P, Marquering W. 2005. The generalized asymmetric dynamic covariance model. *Finance Research Letters* **2**, 67-74.

De Wet WA. 2006. A structural GARCH model: An application on South African data. *Economic Modelling* **23**, 775-791.

Engle RF, Kroner KF. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122-150.

Engle RF, Russell JR. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66**, 1127-1162.

Fazio G. 2007. Extreme interdependence and extreme contagion between emerging markets. *Journal of International Money and Finance* **26**, 1261-1291.

Fleming J, Kirby C, Ostdiek B. 1998. Information and volatility linkages in the stock, bonk and money markets. *Journal of Financial Economics* **49**, 111-137.

Gannon GL. 2004. Simultaneous volatility transmission and spillover effects. School Working Paper Series 2004/10, Deakin University.

Gannon GL. 2005. Simultaneous volatility transmissions and spillover effects: U.S. and Hong Kong stock and futures markets. *International Review of Financial Analysis* **14**, 326-336.

Gannon GL, Choi D. 1998. Structural models: Intra/inter-day volatility transmission and spillover persistence of the HSI, HSIF and S&P500 futures. *International Review of Financial Analysis* **7**, 19-36.

Gourieroux C, Jasiak J. 2001. *Financial Econometrics*. Princeton University Press, Princeton.

Koch PD, Koch TW. 1991. Evolution in dynamic linkages across daily national stock indexes. *Journal of International Money and Finance* **10**, 231-251.

Kroner KF, Ng VK. 1998. Modeling asymmetric comovements of asset returns. *The Review of Financial Studies* **11**, 817-844.

Lee KY 2006. The contemporaneous interactions between the U.S., Japan and Hong Kong stock markets. *Economics Letters* **90**, 21-27.

McNeil AJ, Frey R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**, 271-300.

Rigobon R. 2003. Identification through heteroskedasticity. *The Review of Economics and Statistics* **LXXXV**, 777-792.

Rigobon R, Sack B. 2003. Spillovers across U.S. financial markets. NBER Working Paper 9640.

A Corrected Value-at-Risk Predictor*

Carl Lönnbark
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
Email: carl.lonnbark@econ.umu.se

Abstract

In this note it is argued that the estimation error in Value-at-Risk predictors gives rise to underestimation of portfolio risk. We propose a simple correction and find in an empirical illustration that it is economically relevant.

Key Words: Estimation Error, Finance, GARCH, Prediction, Risk Management.
JEL Classification: C32, C51, C53, G10.

---

*This is a preprint of an article whose final and definitive form will be published in *Applied Economics Letters*. The financial support from the Wallander-Hedelius Foundation is gratefully acknowledged. Kurt Brännäs is thanked for many valuable comments and suggestions. The paper has benefitted from discussions with David Granlund. The author would also like to thank the participants at the 28th annual International Symposium on Forecasting, Nice 2008 and the 2nd International Workshop on Computational and Financial Econometrics, Neuchâtel 2008.

1 Introduction

Value-at-Risk (VaR) has become a standard measure of market risk and it is widely used by financial institutions and their regulators. The VaR is defined as the maximum potential portfolio loss that will not be exceeded with a given probability, or

$$\Pr\left\{\text{portfolio loss} \geq VaR^{1-\alpha}\right\} = \alpha. \quad (1)$$

The Basel Committee on Banking Supervision, which includes governors of the main central banks, imposes on financial institutions such as banks and investment firms to meet capital requirements based on VaR. Accurate VaR estimates are therefore crucially important, and VaR has already received much attention in the literature (see Jorion (2007) for a survey). Although the literature dealing with different modeling issues is large, surprisingly little is written about the uncertainty of VaR predictors. Three studies attempting to quantify the uncertainty are Jorion (1996), Christoffersen and Gonçalves (2005) and Chan, Deng, Peng, and Xia (2007).

The question a practitioner naturally poses is how uncertainty in the VaR affects risk management, i.e. does it in some way change what value to report. Tsay (2005, ch. 7) points out that VaR should be computed using the predictive distribution of returns and should take into account the parameter uncertainty in a properly specified model. The uncertainty arises from two primary sources. The true data generating process is not known, which gives rise to model risk, and the parameters of the hypothesized model must be estimated, which gives rise to estimation risk.

The focus of this paper is on how to incorporate the estimation error in the VaR predictor. In particular, we take a time series model and demonstrate that the implied conventional plug-in VaR predictor does not satisfy eq. (1) asymptotically. In fact, if $VaR^{1-\alpha}$ in eq. (1) is replaced by its predictor, a stochastic variable, the corresponding probability is higher than α, i.e. the portfolio risk is underestimated. This is of course an undesirable feature, but it is relatively straightforward to correct the predictor to give the correct risk measure interpretation. We propose a corrected VaR predictor that accounts for estimation risk. Schaller (2002) discusses along similar lines and suggests an alternative approach. We emphasize that the correction is due to the randomness of the VaR predictor and it is not due to conventional bias, i.e. that the expected value of the VaR predictor is different from the true value. Two studies attempting to correct for conventional bias are Bao and Ullah (2004) and Hartz, Mittnik, and Paolella (2006).

2 VaR and uncertainty

A general multivariate time series model with conditional mean and variance is

$$\mathbf{y}_t = \boldsymbol{\mu}\left(\boldsymbol{\theta}_1, \Psi_{t-1}\right) + \mathbf{H}^{1/2}\left(\boldsymbol{\theta}_2, \Psi_{t-1}\right)\boldsymbol{\varepsilon}_t, \qquad t = 1, ..., T, \tag{2}$$

where Ψ_{t-1} is the information set at $t-1$, $\boldsymbol{\varepsilon}_t \sim i.i.d.\left(\mathbf{0}, \mathbf{I}\right)$ and $\boldsymbol{\mu}$ and $\mathbf{H}$ are, respectively, vector and matrix valued functions. In the sequel we occasionally use the shorthand notation $\boldsymbol{\mu}_t = \boldsymbol{\mu}\left(\boldsymbol{\theta}_1, \Psi_{t-1}\right)$ and $\mathbf{H}_t = \mathbf{H}\left(\boldsymbol{\theta}_2, \Psi_{t-1}\right)$.

Models of this type are usually estimated by maximizing the log-likelihood function $\ln L_T\left(\boldsymbol{\theta}\right) \propto -\frac{1}{2}\sum_t^T [\ln|\mathbf{H}_t| + (\mathbf{y}_t - \boldsymbol{\mu}_t)'\mathbf{H}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t)]$, i.e. assuming $\boldsymbol{\varepsilon}_t \sim n.i.d.\left(\mathbf{0}, \mathbf{I}\right)$. Given some regularity conditions the estimator vector, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2')'$, is asymptotically, normally distributed with the true parameter vector, $\boldsymbol{\theta}_0$, as its mean and covariance matrix $\boldsymbol{\Sigma} = -\left[E\left(\partial^2 \ln L_T\left(\boldsymbol{\theta}_0\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right)\right]^{-1}$.

For a portfolio of assets with returns generated by model (2), the one-period-ahead conditional VaR, $VaR_{T+1}^{1-\alpha}$, satisfies $\Pr(\mathbf{w}_T'\mathbf{y}_{T+1} \leq -VaR_{T+1}^{1-\alpha}|\Psi_T) = \alpha$, where $\mathbf{w}_T$ is a vector of portfolio weights that remains unchanged between T and $T+1$. Assuming normally distributed errors a predictor of the VaR is

$$\widehat{VaR}_{T+1}^{1-\alpha} = -\mathbf{w}_T'\boldsymbol{\mu}\left(\hat{\boldsymbol{\theta}}_1, \Psi_T\right) - \Phi^{-1}\left(\alpha\right)\sqrt{\mathbf{w}_T'\mathbf{H}\left(\hat{\boldsymbol{\theta}}_2, \Psi_T\right)\mathbf{w}_T}. \tag{3}$$

Since the parameters of the underlying model are estimated, $\widehat{VaR}_{T+1}^{1-\alpha}$

is subject to estimation error and is therefore random. It can be decomposed as

$$\widehat{VaR}^{1-\alpha}_{T+1} = VaR^{1-\alpha}_{T+1} + e_{T+1} + b_{T+1},$$

where e_{T+1} accounts for sampling variation and has zero mean and variance $\sigma^2_{VaR,T+1}$. The finite sample bias of the predictor is denoted b_{T+1}. In a related study, Hansen (2006) showed that asymptotically $\sqrt{T}e_{T+1} \xrightarrow{d} N(0, T\sigma^2_{VaR,T+1})$, where $\sigma^2_{VaR,T+1} = (\partial VaR^{1-\alpha}_{T+1}/\partial\boldsymbol{\theta}')\boldsymbol{\Sigma}(\partial VaR^{1-\alpha}_{T+1}/\partial\boldsymbol{\theta})$. Due to e_{T+1} and b_{T+1}, $\widehat{VaR}^{1-\alpha}_{T+1}$ satisfies

$$\Pr\left\{\mathbf{w}'_T\mathbf{y}_{T+1} \leq -\widehat{VaR}^{1-\alpha}_{T+1}|\Psi_T\right\} = \alpha^*,$$

where α^* need not equal α.

Now, introduce a correction term c_{T+1} such that $\Pr\{\mathbf{w}'_T\mathbf{y}_{T+1} \leq -(VaR^{1-\alpha}_{T+1} + e_{T+1} + b_{T+1} + c_{T+1})|\Psi_T\} = \alpha$. Assume that $\mathbf{y}_{T+1}$ and e_{T+1} are independent[1] and we have that

$$\Pr\left\{\frac{\mathbf{w}'_T\mathbf{y}_{T+1} + e_{T+1} - \mathbf{w}'_T\boldsymbol{\mu}_{T+1}}{\sqrt{\mathbf{w}'_T\mathbf{H}_{T+1}\mathbf{w}_T + \sigma^2_{VaR,T+1}}} \leq \right.$$
$$\left. \frac{-\left(VaR^{1-\alpha}_{T+1} + b_{T+1} + c_{T+1}\right) - \mathbf{w}'_T\boldsymbol{\mu}_{T+1}}{\sqrt{\mathbf{w}'_T\mathbf{H}_{T+1}\mathbf{w}_T + \sigma^2_{VaR,T+1}}}|\Psi_T\right\} = \alpha.$$

By the normality of e_{T+1} and $\mathbf{y}_{T+1}$ the correction can be obtained as

$$c_{T+1} = -\Phi^{-1}(\alpha)\left[\sqrt{\mathbf{w}'_T\mathbf{H}_{T+1}\mathbf{w}_T + \sigma^2_{VaR,T+1}} - \sqrt{\mathbf{w}'_T\mathbf{H}_{T+1}\mathbf{w}_T}\right] - b_{T+1}.$$

Due to b_{T+1}, the correction may in small samples be either positive or negative. Asymptotically, however, b_{T+1} is zero, c_{T+1} is positive, and then $\alpha^* > \alpha$. Now, add the estimator of the correction to the conventional predictor, $\widehat{VaR}^{1-\alpha}_{T+1}$, and the corrected VaR predictor becomes

$$\widehat{CVaR}^{1-\alpha}_{T+1} = -\mathbf{w}'_T\hat{\boldsymbol{\mu}}_{T+1} - \Phi^{-1}(\alpha)\sqrt{\mathbf{w}'_T\hat{\mathbf{H}}^{1/2}_{T+1}\mathbf{w}_T + \hat{\sigma}^2_{VaR,T+1}}.$$

[1]In the simulation exercise in Section 3, the covariance between the $\mathbf{y}_{T+1}$ and the e_{T+1} series was close to zero for all four sample sizes.

Table 1: Statistics for the corrections in the numerical illustration. S.D. is the standard deviation. The z's are the test statistics for the hypotheses $\alpha^* = 0.01$ and $\alpha^c = 0.01$, respectively.

T	Mean	S.D.	Skewness	Kurtosis	z_{α^*}	z_{α^c}
250	0.061	0.063	5.007	60.236	3.909	-0.318
500	0.028	0.027	4.272	34.852	1.716	-0.222
1000	0.013	0.013	4.196	31.151	0.254	-0.636
2500	0.005	0.005	4.304	34.284	0.032	-0.254

3 A numerical illustration

To illustrate the properties of the correction we conduct a small simulation experiment with data generated according to a GARCH$(1,1)$ model: $y_t = \sqrt{h_t}\varepsilon_t$, with $h_t = 40/252 + 0.1y_{t-1}^2 + 0.8h_{t-1}$ and $\varepsilon_t \sim n.i.d.\,(0,1)$. We consider samples of sizes $250, 500$, 1000, and 2500 observations, $\alpha = 0.01$ and results are in each case based on 100 000 replications. The fractions of exceedences for the conventional (α^*) and the corrected VaR predictor (α^c) are computed and the hypotheses $\alpha^* = \alpha$ and $\alpha^c = \alpha$ are tested against the one-sided alternatives $\alpha^* > \alpha$ and $\alpha^c > \alpha$, respectively. Table 1 gives some statistics for the estimated corrections and the z-statistics for the tests.

As expected the mean and the variance of the correction decreases with the sample size. The hypothesis $\alpha^* = \alpha$ is rejected at the 5% level for the sample sizes 250 and 500, while the hypothesis $\alpha^c = \alpha$ is not rejected for any of the four sample sizes. That is, the correction matters statistically for the two smaller sample sizes and it appears that it does the job of taking the estimation error into account.

4 An empirical illustration

VaR corrections are obtained for the three major stock market indices: FTSE 100 of UK, Nikkei 225 of Japan, and S&P 100 of USA. Five years of daily index data were downloaded from DataStream and returns were calculated as $y_t = 100 \times \log(I_t/I_{t-1})$, where I_t is the value of the index at t. The sample covers February 6, 2003 to February 7, 2008, for a total of 1304 observations. We consider $\alpha = 0.01$ and the predictor for VaR at $t+1$ is based on observations $t-653$ to t, $t = 654, ..., 1304$. VaR's are predicted for the final half of the sample, and are based on re-estimated GARCH$(1,1)$ models with constant means. Figure 1 gives the estimated corrections in percentage points.

The corrections exhibit time variation and vary between 0.003 and 0.140 for the FTSE 100 index, 0.003 and 0.069 for Nikkei 225 and 0.002 and 0.183 for S&P 100. The average corrections are 0.016, 0.014 and 0.015. These small numbers must be converted into monetary units to give a fair picture. For example, a correction of 0.05 for a portfolio with 100 billion dollars worth of assets is 50 million dollars on a daily basis.

The few relatively large corrections are due to outliers and highlight the sensitivity of both the Quasi-Maximum Likelihood and the $\sigma^2_{VaR,T+1}$ estimators to extreme observations.

5 Conclusion

This note argued that the estimation error in VaR predictors gives rise to underestimation of portfolio risk. We introduced an approach to correcting a predictor to account for the estimation error, and in an empirical illustration we found that the correction is of economic relevance. The proposed correction hinges on the normality of both the VaR estimator and the returns and does not apply directly to cases with non-normally distributed returns. Adapting the proposed approach to other distributions is in principle straightforward, though.

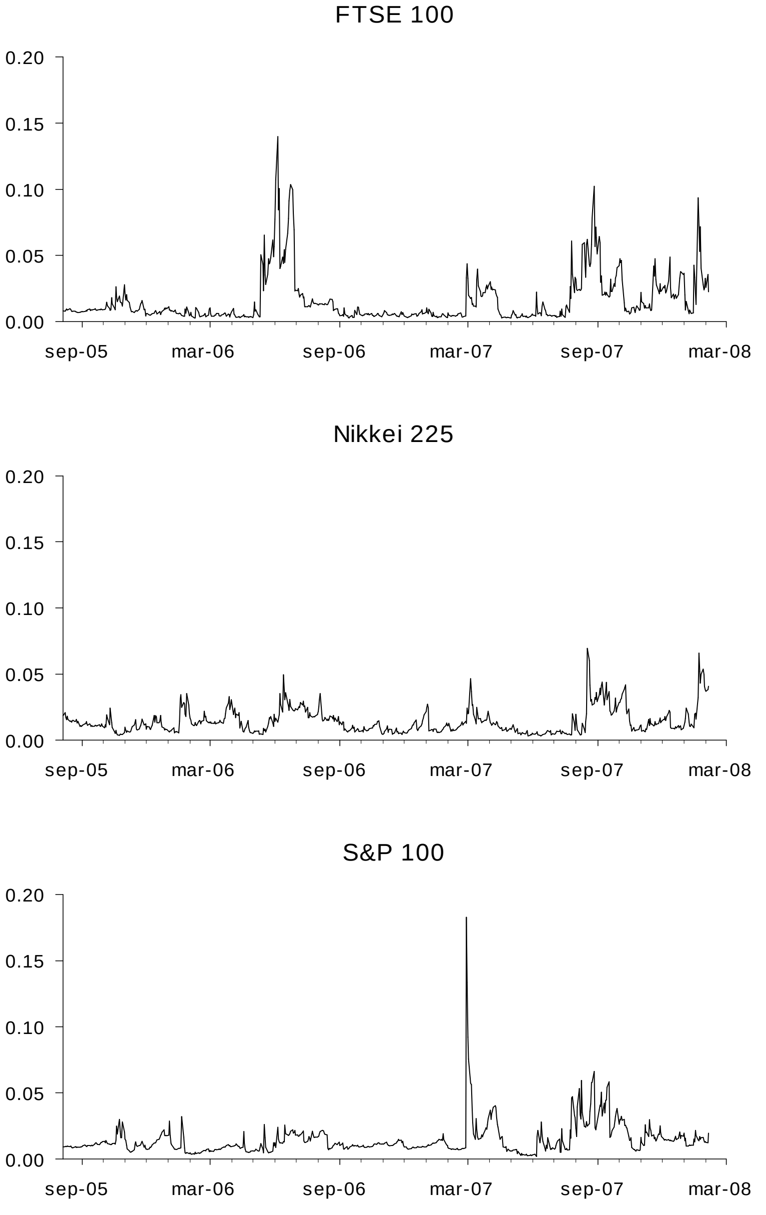

Figure 1: Corrections in percentage points.

References

Bao, Y., Ullah, A., 2004. Bias of a Value-at-Risk estimator. Finance Research Letters 1, 241–249.

Chan, N. H., Deng, S.-J., Peng, L., Xia, Z., 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics 137, 556–576.

Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. Journal of Risk 7, 1–28.

Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. Journal of Econometrics 135, 377–398.

Hartz, C., Mittnik, S., Paolella, M., 2006. Accurate value-at-risk forecasting based on the normal-GARCH model. Computational Statistics & Data Analysis 51, 2295–2312.

Jorion, P., 1996. Risk^2: Measuring the Risk in Value at Risk. Financial Analysts Journal 52, 47–56.

Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill, New York.

Schaller, P., 2002. Uncertainty of parameter estimates in VAR calculations. Bank Austria Working Paper BA-RISK-20020419.

Tsay, R. S., 2005. Analysis of Financial Time Series, 2nd Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

Uncertainty of Multiple Period Risk Measures*

Carl Lönnbark
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
E-mail: carl.lonnbark@econ.umu.se

Abstract

In general, the properties of the conditional distribution of multiple period returns do not follow easily from the one-period data generating process. This renders computation of the Value at Risk and the Expected Shortfall for multiple period returns a non-trivial task. In this paper we consider some approximation approaches to computing these measures. Based on the result of a simulation experiment we conclude that among the approaches studied the one based on assuming a skewed t distribution for the multiple period returns and that based on simulations were the best. We also found that the uncertainty due to the estimation error can be quite accurately estimated employing the delta method. In an empirical illustration we computed five day Value at Risk's for the S&P 500 index. The approaches performed about equally well.

Key Words: Asymmetry, Estimation Error, Finance, GJR-GARCH, Prediction, Risk Management.
JEL Classification: C16, C46, C52, C53, C63, G10.

*The financial support from the Jan Wallander and Tom Hedelius Foundation and the Tore Browaldh Foundation is gratefully acknowledged. Kurt Brännäs and Johan Lyhagen are thanked for many valuable comments and suggestions.

1 Introduction

The focus of this paper is on predicting the risk for multiple period asset returns. An important example when this is of interest is for the market-risk charge of the Basel Committee on Banking Supervision (Basel), that is based on an horizon of 10 trading days. The market risk is defined as the risk of adverse movements in the prices of the assets in the portfolio and the measure underlying the market risk charge is the Value at Risk (VaR) (defined below). Basel allows financial institutions to compute the 10 day VaR by multiplying the one day VaR by the square root of 10. However, it is well known (e.g., Diebold, Hickman, Inoue, and Schuermann, 1997) that this approach (Root-k) may give very erroneous VaR's and alternative approaches are thus called for.

When it comes to predicting more than one period ahead there are two approaches: The direct approach specifies a model for the relevant horizon, e.g., 10 days, directly, whereas the iterating approach iterates on a model specified for a shorter horizon, e.g., one day, to obtain the multiple period predictions. The first approach may be more robust to misspecification, while the latter may produce more efficient parameter estimates (e.g., Marcellino, Stock, and Watson, 2006; Pesaran, Pick, and Timmermann, 2009). The recommendation put forth by Diebold et al. (1997) is to use the direct approach for risk predictions. Taylor (1999, 2000) propose a regression quantile approach that may be viewed as a combination of the two. In practise, the computed risk measures are subject to estimation error. Assume for example that we wish to predict the risk of an asset for a 10 day horizon and that we have two years of daily return data. For the iterating approach we would typically specify a model for the daily returns and base the prediction on the full sample of approximately 500 observations. For the direct approach on the other hand we would have only 50 observations, which may not be enough for producing a reliable prediction. We view this as a valid concern and focus here on the iterating approach. Of course, an important underlying question that we neglect here is that of whether the properties of the return distribution can be considered predictable for a particular horizon (see Christoffersen and Diebold, 2000, for a discussion on volatility predictability).

As measures of (market) risk we consider VaR and the Expected Shortfall (ES). The VaR has become the standard measure of market risk and it is commonly employed by financial institutions and their regulators. The VaR has already received much attention in the literature (see Jorion, 2007, for a survey) and it is defined as the maximum potential loss over a given horizon that will not be exceeded with a given probability, or

$$\Pr\left\{\text{portfolio loss} \geq VaR^{1-\alpha}\right\} = \alpha.$$

The probability $1-\alpha$ is commonly referred to as the confidence level of the VaR. The attractive feature of the VaR is that it summarizes the properties of the return distribution into an easily interpreted number. However, it does not tell the risk manager anything about the size of the loss when disaster strikes. A measure that does exactly that is the ES. It is defined as

$$ES^{1-\alpha} = E\left(\text{portfolio loss} \mid \text{portfolio loss} \geq VaR^{1-\alpha}\right).$$

Suppose now that the risk manager wants to assess the k-period risk of the portfolio and decides to employ the iterating approach within the popular GARCH framework of Engle (1982) and Bollerslev (1986). A problem that arises is then that the properties of the multiple period return distribution may not follow easily from the one-period model. For example, even though the multiple period conditional variance implied from a one-period GARCH model with normal innovations is tractable, less so is the distribution of the corresponding innovation (Boudoukh, Richardson, and Whitelaw, 1997). Brummelhuis and Guégan (2005) provide a theoretical discussion on the matter. In particular, they show that the Root-k rule may fail severely for small values on α (see also Brummelhuis and Kaufmann, 2007).

Two alternative approaches are to compute the measures either by simulation (cf. McNeil and Frey, 2000) or to consider some analytic approximation. The former computes the measures as empirical counterparts for multiple period returns simulated from the one-period model.

Assuming that the true parameters of the one-period model are known, the simulation approach can give measures arbitrarily close to the true ones. We will discuss two analytical approximations. The first one uses a Gram-Charlier expansion of the conditional density of the multiple period returns. The second one was proposed by Wong and So (2003, 2007) in related studies. It consists of specifying a conditional distribution for the multiple period returns and of obtaining the parameters of that distribution by matching its moments to the theoretical ones implied by the one-period model. The obvious benefit of using analytic approximations is that they require less computer time. In Cotter (2007) an approach based on extreme value theory is proposed. It performed poorly in simulations, though, and we do not consider it here.

As noted above, an additional source of uncertainty of the risk predictors arises from the fact that the parameters of the underlying model are unknown, which gives rise to estimation error. We also pay attention to this source of error, which is not done in Wong and So (2003). Note that this uncertainty comes in in two places for the simulation based predictor. Not only in estimating the parameters of the underlying model, but also in the second step when the measures are obtained from the simulated returns.

The uncertainty in risk prediction should be of concern to risk managers. Surprisingly little work has been done on it though and the predictions are often reported as if they were true constants. For example Lan, Hu, and Johnson (2007) report that the research on the uncertainty of VaR only amounts to about 2.5 percent of the VaR literature. One study that recognizes that VaR and ES predictors are subject to uncertainty is Christoffersen and Gonçalves (2005), who use resampling techniques to study the uncertainty of VaR and ES predictors in a GARCH framework. The obvious disadvantage of their method is that it is time consuming since it amounts to repeated estimation of a possibly complicated model. Analytical expressions (when sufficiently accurate) to quantify the uncertainty are obviously preferred. For this purpose Chan, Deng, Peng, and Xia (2007) and others consider the conventional delta method, which is done here as well.

The paper is organized as follows. In Section 2 the approaches to computing the multiple period VaR and ES are introduced. In Section 3 we discuss how to quantify the uncertainty due to the estimation error. An example is given in Section 4, where analytical results are given for the asymmetric GARCH (GJR-GARCH) model of Glosten, Jagannathan, and Runkle (1993). Section 5 contains a simulation study of the predictors obtained from the GJR-GARCH. In Section 6 an empirical illustration for the S&P 500 index is included. The final section concludes.

2 Multiple period VaR and ES

Denote by $\mathbf{w} = (w_1, ..., w_M)'$ the time invariant vector of portfolio weights between T and $T+k$. The log-return (return) between T and $T+k$ for the portfolio is approximately $\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'(\mathbf{y}_{T+1} + ... + \mathbf{y}_{T+k})$, where $\mathbf{y}_{T+l} = (y_{1,T+l}, ..., y_{M,T+l})'$, $l = 1, ..., k$, is a M-dimensional vector of one-period returns. Denote by Ψ_T the information set at time T and let the vector $\boldsymbol{\theta}$ contain the parameters governing the data generating process with $\boldsymbol{\theta}_0$ denoting true values. In practise, the information available to the risk manager is some realization of the partition, $\mathcal{F}_{t_0,T} = (\mathbf{x}_{t_0}, ..., \mathbf{x}_T)$, of Ψ_T and where $\mathbf{x}_t$, $t = t_0, ..., T$, typically contains past asset returns. A realization of the random partition, $\mathcal{F}_{t_0,T}$, is denoted by $\mathcal{F}_{t_0,T}$. Denote by $f_{T,k}(\cdot)$ and $F_{T,k}(\cdot)$ the density function (*pdf*) and distribution function (*cdf*) of $\mathbf{w}'\mathbf{Y}_{T,k}$ conditional on Ψ_T. Also, let $\boldsymbol{\mu}_{T,k}$ be the vector valued conditional mean function and $\mathbf{H}_{T,k}$ the matrix valued conditional variance-covariance function of $\mathbf{Y}_{T,k}$. We will assume that it is possible to obtain the exact forms of these conditional moments for all k.

Now, assume that the vector process, $\mathbf{y}_t$, of the asset returns started in the infinite past and that it is generated in discrete time up through, at least, $T+k$ by

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{H}_t^* \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{\varepsilon}_t$ has mean $\mathbf{0}$ and the identity matrix, $\mathbf{I}$, as its variance-covariance matrix conditional on the information available at $t-1$. Then, $\boldsymbol{\mu}_t$ is the conditional mean of $\mathbf{y}_t$, whereas $\mathbf{H}_t = \mathbf{H}_t^* \mathbf{H}_t^{*\prime}$ is the conditional variance-covariance matrix.

The conditional VaR for the period T to $T+k$ portfolio return satisfies

$$P\left(\mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha}|\Psi_T\right) = \int_{-\infty}^{-VaR_{T,k}^{1-\alpha}} f_{T,k}\left(y\right)dy = \alpha. \qquad (2)$$

The associated conditional ES is defined as

$$\begin{aligned} ES_{T,k}^{1-\alpha} &= -E_T\left(\mathbf{w}'\mathbf{Y}_{T,k} \mid \mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha}\right) \qquad (3) \\ &= -\frac{1}{\alpha}\int_{-\infty}^{-VaR_{T,k}^{1-\alpha}} y f_{T,k}\left(y\right)dy, \end{aligned}$$

where $E_T\left(\cdot\right)$ is shorthand for expectation conditional on Ψ_T. The minus signs in (2) and (3) stem from the convention of reporting VaR and ES as positive numbers.

For $k=1$, $VaR_{T,1}^{1-\alpha}$ and $ES_{T,1}^{1-\alpha}$ can (in principle) be obtained directly from (1) along with a distributional assumption on $\boldsymbol{\varepsilon}_{T+1}$. Although complications may arise in this case as well we choose here to focus on the case when $k>1$. The further issue is then one of temporal aggregation and our point of departure is that it is not possible to obtain $VaR_{T,k}^{1-\alpha}$ and $ES_{T,k}^{1-\alpha}$ analytically and that we have to resort to some approximations $\widetilde{VaR}_{T,k}^{1-\alpha}$ and $\widetilde{ES}_{T,k}^{1-\alpha}$. We consider three such approaches. One is simulation based and targets the measures directly, whereas the other two are analytical approximations and start from an approximation to a zero mean and unit variance random variable, $\varepsilon_{T,k}$.

Denote by $VaR_{T,k}^{S,1-\alpha}$ and $ES_{T,k}^{S,1-\alpha}$ the values of $\widetilde{VaR}_{T,k}^{1-\alpha}$ and $\widetilde{ES}_{T,k}^{1-\alpha}$ computed by the simulation approach. To explain the approach, we first assume that observations are available up through T. We then simulate returns $\mathbf{y}_{T+1}^r$, $\mathbf{y}_{T+2}^r$, ..., $\mathbf{y}_{T+k}^r$, $r=1,...,R$, from the model (1) and compute the k-period portfolio returns $\mathbf{w}'\mathbf{Y}_{T,k}^r$, $r=1,...,R$. The VaR is obtained as the αth empirical quantile of the simulated portfolio returns, or

$$VaR_{T,k}^{S,1-\alpha} = -\left(\mathbf{w}'\mathbf{Y}_{T,k}\right)_{(\alpha R+1)},$$

where $(\mathbf{w}'\mathbf{Y}_{T,k})_{(r)}$ is the rth order statistic of the simulated returns. The corresponding ES is given by

$$ES^{S,1-\alpha}_{T,k} = -\frac{\sum_{r=1}^{\alpha R+1}(\mathbf{w}'\mathbf{Y}_{T,k})_{(r)}}{\alpha R+1}.$$

Note that $\mathbf{w}'\mathbf{Y}^r_{T,k}$ is iid and it is well-known that the resulting estimators are consistent. Given R though, one may of course argue that more efficient related estimators based on kernel functions exist. Chen and Tang (2005) and Chen (2008) found that, for the kernel estimator proposed by Scaillet (2004), this is the case for VaR but not necessarily for ES. Note however that R is at our discretion and extra precision comes at a small marginal cost for models within a reasonable degree of complexity.

For the analytical approaches we first assume that the k-period portfolio return, $\mathbf{w}'\mathbf{Y}_{T,k}$, admits the scale-location representation

$$\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'\boldsymbol{\mu}_{T,k} + \varepsilon_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{4}$$

where $\varepsilon_{T,k}$ has zero mean, unit variance, conditional third moment, $s_{T,k}$, conditional fourth moment, $k_{T,k}$, and conditional density function

$$g_{T,k}(\varepsilon) = \sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}\, f_{T,k}(\boldsymbol{\mu}_{T,k} + \varepsilon\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}).$$

From (4) we then have that

$$\begin{aligned} P\left(\mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR^{1-\alpha}_{T,k}|\Psi_T\right) &= P\left(\mathbf{w}'(\mathbf{Y}_{T,k} - \boldsymbol{\mu}_{T,k})/\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}\right. \\ &\left.\leq (-VaR^{1-\alpha}_{T,k} - \mathbf{w}'\boldsymbol{\mu}_{T,k})/\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}|\Psi_T\right) = P\left(\varepsilon_{T,k} \leq q^{\alpha}_{T,k}|\Psi_T\right), \end{aligned}$$

where $q^{\alpha}_{T,k}$ solves $\alpha = \int_{-\infty}^{q^{\alpha}_{T,k}} g_{T,k}(\varepsilon)\, d\varepsilon$. The conditional portfolio VaR is then given by

$$VaR^{1-\alpha}_{T,k} = -\mathbf{w}'\boldsymbol{\mu}_{T,k} - q^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}.$$

The conditional ES of the portfolio is

$$ES^{1-\alpha}_{T,k} = -\mathbf{w}'_T\boldsymbol{\mu}_{T,k} - e^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}},$$

where $e^{\alpha}_{T,k} = E_T(\varepsilon_{T,k} \mid \varepsilon_{T,k} \leq q^{\alpha}_{T,k})$. We previously assumed that it was possible to obtain the exact analytical forms of $\boldsymbol{\mu}_{T,k}$ and $\mathbf{H}_{T,k}$. The problem is then one of approximating the density $g_{T,k}(\cdot)$. Denote this approximation by $\tilde{g}_{T,k}(\cdot)$ and the associated VaR and ES are then

$$\widetilde{VaR}^{1-\alpha}_{T,k} = -\mathbf{w}'\boldsymbol{\mu}_{T,k} - \tilde{q}^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{5}$$

$$\widetilde{ES}^{1-\alpha}_{T,k} = -\mathbf{w}'\boldsymbol{\mu}_{T,k} - \tilde{e}^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{6}$$

where $\tilde{q}^{\alpha}_{T,k}$ satisfies $\alpha = \int_{-\infty}^{\tilde{q}^{\alpha}_{T,k}} \tilde{g}_{T,k}(\varepsilon)\, d\varepsilon$ and $\tilde{e}^{\alpha}_{T,k} = \widetilde{E}_T(\varepsilon_{T,k} \mid \varepsilon_{T,k} \leq \tilde{q}^{\alpha}_{T,k})$. Note that $\widetilde{E}_T$ is the expectation operator with respect to $\tilde{g}_{T,k}(\cdot)$.

Our first analytical approximation employs an expansion of $g_{T,k}(\cdot)$ allowing for skewness and excess kurtosis. We assume that $g_{T,k}(\cdot)$ admits the Gram-Charlier Type A expansion

$$g_{T,k}(\varepsilon) = \sum_{i=0}^{\infty} c_i H_i(\varepsilon)\,\phi(\varepsilon), \tag{7}$$

where the constants, c_i, are functions of the conditional moments of $\varepsilon_{T,k}$, $H_i(\cdot)$ are the Hermite polynomials and $\phi(\cdot)$ is the standard normal *pdf*. The sum in (7) is usually truncated at a small value of i. Jondeau and Rockinger (2001) identify the versions typically adopted in the literature to be the Edgeworth expansion and the Gram-Charlier expansion. The latter is given by

$$\tilde{g}_{T,k}(\varepsilon) = [1 + \frac{s_{T,k}}{6} H_3(\varepsilon) + \frac{k_{T,k} - 3}{24} H_4(\varepsilon)]\phi(\varepsilon). \tag{8}$$

The Edgeworth expansion adds the term $s^2_{T,k} H_6(\varepsilon)/72$ to the expression inside the brackets in (8). Barton and Dennis (1952) show that the region of $(s_{T,k}, k_{T,k})$-pairs guaranteeing positive values is larger for the Gram-Charlier expansion, and for that reason, Jondeau and Rockinger (2001) focus on the latter and so do we.

The αth quantile implied by the Gram-Charlier density (8) is given by the Cornish-Fisher expansion (Cornish and Fisher, 1938; see also Baillie and Bollerslev, 1992, for a related use)

$$\begin{aligned}\tilde{q}^{\alpha}_{T,k} &= \Phi^{-1}_{\alpha} + \frac{s_{T,k}}{6}[\left(\Phi^{-1}_{\alpha}\right)^2 - 1] \\ &\quad + \frac{k_{T,k}-3}{24}[\left(\Phi^{-1}_{\alpha}\right)^3 - 3\Phi^{-1}_{\alpha}] - \frac{s^2_{T,k}}{36}[2\left(\Phi^{-1}_{\alpha}\right)^3 - 5\Phi^{-1}_{\alpha}].\end{aligned}$$

The third, $s_{T,k}$, and the fourth, $k_{T,k}$, conditional moments of $\varepsilon_{T,k}$ are derived from the one-period model. Christoffersen and Gonçalves (2005) propose a corresponding $\tilde{e}^{\alpha}_{T,k}$. Giamouridis (2006) correctly argues that their expression is incorrect and propose

$$\tilde{e}^{\alpha}_{T,k} = -\frac{\phi_{\tilde{q}^{\alpha}_{T,k}}}{\alpha}\left\{1 + \frac{s_{T,k}}{6}\left(\tilde{q}^{\alpha}_{T,k}\right)^3 + \frac{k_{T,k}-3}{24}\left[\left(\tilde{q}^{\alpha}_{T,k}\right)^4 - 2\left(\tilde{q}^{\alpha}_{T,k}\right)^2 - 1\right]\right\}.$$

The VaR and the ES are obtained by plugging the expressions above into (5) and (6), respectively. We denote the resulting approximations by $VaR^{GC,1-\alpha}_{T,k}$ and $ES^{GC,1-\alpha}_{T,k}$.

Alternatively, Wong and So (2003) assume a distribution for $\varepsilon_{T,k}$ and obtaining the parameters of that distribution involves matching the third and the fourth moments to the corresponding ones implied by the one-period model. We denote the resulting approximations by $VaR^{WS,1-\alpha}_{T,k}$ and $ES^{WS,1-\alpha}_{T,k}$.

For comparison we also include the Root-k approach. The k-period VaR and the ES are then simply approximated by

$$\begin{aligned}VaR^{Rk,1-\alpha}_{T,k} &= \sqrt{k}VaR^{1-\alpha}_{T,1}, \\ ES^{Rk,1-\alpha}_{T,k} &= \sqrt{k}ES^{1-\alpha}_{T,1}.\end{aligned}$$

3 Estimation error

The traditional estimator of the parameter vector, $\boldsymbol{\theta}$, in model (1) has over the years been (conditional) maximum likelihood with a normality assumption on $\boldsymbol{\varepsilon}_t$, i.e.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{L_T\left(\boldsymbol{\theta}\right)\right.$$

$$\propto -\frac{1}{2} \sum_{t=t_0+s}^{T} \left[\ln |\mathbf{H}_t| + (\mathbf{y}_t - \boldsymbol{\mu}_t)' \mathbf{H}_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t) \right] \Bigg\} . \tag{9}$$

Here, s is determined by the number of lags in $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$. Given some regularity conditions the estimator, $\hat{\boldsymbol{\theta}}$, is asymptotically, normally distributed with the true parameter vector, $\boldsymbol{\theta}_0$, as its mean and with variance-covariance matrix $\boldsymbol{\Sigma}$, which may be consistently estimated by $T^{-2}[\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}']$ or $\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. As shown by Bollerslev and Wooldridge (1992) and others, the estimator (9) remains consistent and asymptotically normal even if the distribution of $\boldsymbol{\varepsilon}_t$ is non-normal. The estimator is then known as the Quasi-Maximum Likelihood (QML) estimator and we would use the robust sandwich form as the estimator of $\boldsymbol{\Sigma}$, i.e. $(\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}(\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}')(\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}$.

For all four approaches the approximated risk measures are functions of the parameters $\boldsymbol{\theta}$. Therefore, the measures are not only subject to an approximation error, but also to the estimation error in $\hat{\boldsymbol{\theta}}$. In the first approach this shows up in the simulations as they are made from the model (1) under $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The other three predictors are obtained by plugging the estimator, $\hat{\boldsymbol{\theta}}$, into (5) and (6) to obtain

$$\begin{aligned} \widehat{\widetilde{VaR}}{}_{T,k}^{1-\alpha} &= -\mathbf{w}'\hat{\boldsymbol{\mu}}_{T,k} - \widehat{\tilde{q}}{}_{T,k}^{\alpha}\sqrt{\mathbf{w}'\hat{\mathbf{H}}_{T,k}\mathbf{w}} \\ \widehat{\widetilde{ES}}{}_{T,k}^{1-\alpha} &= -\mathbf{w}'\hat{\boldsymbol{\mu}}_{T,k} - \widehat{\tilde{e}}{}_{T,k}^{\alpha}\sqrt{\mathbf{w}'\hat{\mathbf{H}}_{T,k}\mathbf{w}}. \end{aligned}$$

Early attempts (see Schmidt, 1974) to quantify the effect on prediction of errors in parameters relied on the asymptotic distribution of the parameter estimator assumed to be independent of the conditioning information. In the notation set out in the beginning of Section 2 the predictors are functions of $\Psi_{t_0,T}$ both directly and indirectly through $\hat{\boldsymbol{\theta}}$. Denote this (continuous) function by $u\left(\Psi_{t_0,T}, \hat{\boldsymbol{\theta}}(\Psi_{t_0,T})\right)$. The approach then amounts to conditioning the first argument of $u(\cdot)$ on a realization $\bar{\Psi}_{t_0,T}$ and viewing randomness to arise through the random $\Psi_{t_0,T}$ in the second argument. This approach now appears to be the convention (see Kaibila and He, 2004, for a recent discussion) and is chosen in this paper as well. In a related study Hansen (2006) takes this route and shows that, for $k = 1$, $\sqrt{T^*}\left(\widehat{VaR}_{T,1}^{1-\alpha} - VaR_{T,1}^{1-\alpha}\right) \xrightarrow{d} N\left(0, T^*\sigma^2_{VaR,T,1}\right)$, where $T^* = T - (t_0 + s)$, $\sigma^2_{VaR,T,1} = \partial VaR_{T,1}^{1-\alpha}/\partial\boldsymbol{\theta}'\,\boldsymbol{\Sigma}\,\partial VaR_{T,1}^{1-\alpha}/\partial\boldsymbol{\theta}$ and where the limit is for $t_0 \to -\infty$. This approach is directly applicable to the analytical approximations approach. They are all functions of the estimator, $\hat{\boldsymbol{\theta}}$, and the information set, $\Psi_{t_0,T}$. By the same logic as above we have that

$$\sqrt{T^*}\left[u\left(\bar{\Psi}_{t_0,T}, \hat{\boldsymbol{\theta}}\right) - u\left(\bar{\Psi}_{t_0,T}, \boldsymbol{\theta}_0\right)\right] \xrightarrow{d} N\left(0, T^*\sigma_u^2\right),$$

where $\sigma_u^2 = \partial u/\partial\boldsymbol{\theta}'\boldsymbol{\Sigma}\partial u/\partial\boldsymbol{\theta}$ and note that it is a function of $\bar{\Psi}_{t_0,T}$. Explicitly, the variance expressions for the VaR and ES approximations are, respectively, given by

$$\begin{aligned}
\sigma^2_{VaR,T,k} &= \frac{\partial\widetilde{VaR}_{T,k}^{1-\alpha}}{\partial\boldsymbol{\theta}'}\,\boldsymbol{\Sigma}\,\frac{\partial\widetilde{VaR}_{T,k}^{1-\alpha}}{\partial\boldsymbol{\theta}} \\
\sigma^2_{ES,T,k} &= \frac{\partial\widetilde{ES}_{T,k}^{1-\alpha}}{\partial\boldsymbol{\theta}'}\,\boldsymbol{\Sigma}\,\frac{\partial\widetilde{ES}_{T,k}^{1-\alpha}}{\partial\boldsymbol{\theta}},
\end{aligned}$$

where $\partial\widetilde{VaR}_{T,k}^{1-\alpha}/\partial\boldsymbol{\theta} = -\mathbf{w}'\partial\boldsymbol{\mu}_{T,k}/\partial\boldsymbol{\theta} - \partial\tilde{q}_{T,k}^{\alpha}/\partial\boldsymbol{\theta}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}} - \tilde{q}_{T,k}^{\alpha}\mathbf{w}'$ $\partial\mathbf{H}_{T,k}/\partial\boldsymbol{\theta}\mathbf{w}/(2\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}})$ and $\partial\widetilde{ES}_{T,k}^{1-\alpha}/\partial\boldsymbol{\theta} = -\mathbf{w}'\partial\boldsymbol{\mu}_{T,k}/\partial\boldsymbol{\theta} - \partial\tilde{e}_{T,k}^{\alpha}/$ $\partial\boldsymbol{\theta}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}} - \tilde{e}_{T,k}^{\alpha}\mathbf{w}'\partial\mathbf{H}_{T,k}/\partial\boldsymbol{\theta}\mathbf{w}/(2\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}})$. In practise, estimators of the derivatives are obtained by plugging in $\hat{\boldsymbol{\theta}}$.

Regarding the uncertainty of the simulation based predictor we first recognize that it is a two-step procedure. The first step consists of estimating the model based on the available observations, whereas the predictors in the second step are obtained based on simulated returns from the estimated model. Hence, the estimation uncertainty comes from two sources. Now, for notational convenience drop the time indices on the *pdf* and the *cdf* of the k-period portfolio return and extend the functions to $f(\cdot;\boldsymbol{\theta})$ and $F(\cdot;\boldsymbol{\theta})$ to indicate the value of the parameter. Also, let $v_{\hat{\boldsymbol{\theta}}}$ and $e_{\hat{\boldsymbol{\theta}}}$ (not to be confused with $e_{T,k}^{\alpha}$ above) denote the true $VaR_{T,k}^{1-\alpha}$ and $ES_{T,k}^{1-\alpha}$ under the parameterization $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Now, it is possible to show (see Manistre and Hancock, 2005, and references therein) that conditional on $\hat{\boldsymbol{\theta}}$

$$\hat{v}_{\hat{\boldsymbol{\theta}}} \overset{asy}{\sim} N\left(v_{\hat{\boldsymbol{\theta}}}, V_{\hat{\boldsymbol{\theta}}}^{v}\right) \tag{10}$$

$$\hat{e}_{\hat{\boldsymbol{\theta}}} \overset{asy}{\sim} N\left(e_{\hat{\boldsymbol{\theta}}}, V^e_{\hat{\boldsymbol{\theta}}}\right), \tag{11}$$

where $V^v_{\hat{\boldsymbol{\theta}}} = \alpha(1-\alpha)/(f(v_{\hat{\boldsymbol{\theta}}};\hat{\theta})^2 R)$ and $V^e_{\hat{\boldsymbol{\theta}}} = [V(Y_{T,k} \mid Y_{T,k} < v_{\hat{\boldsymbol{\theta}}}) + (1-\alpha)(e_{\hat{\boldsymbol{\theta}}} - v_{\hat{\boldsymbol{\theta}}})^2]/(R\alpha)$. Of course, these variances do not recognize that $\hat{\boldsymbol{\theta}}$ is random. To derive such expressions we use the variance decomposition formula and take a first order expansion around $\boldsymbol{\theta}_0$. Ignoring higher order terms we have for $\hat{v}_{\boldsymbol{\theta}}$ that

$$\begin{aligned} V(\hat{v}_{\hat{\boldsymbol{\theta}}}) &= E[V^v_{\hat{\boldsymbol{\theta}}}] + V[v_{\hat{\boldsymbol{\theta}}}] \\ &= E[V^v_{\boldsymbol{\theta}_0} + \frac{\partial V^v_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] + V[v_{\boldsymbol{\theta}_0} + \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] \\ &= V^v_{\boldsymbol{\theta}_0} + \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}'}\, \boldsymbol{\Sigma}\, \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}, \end{aligned}$$

where the expectation and the variance are taken over $\hat{\theta}$, and where the first approximation is motivated by (10) and the second and third ones by the asymptotic properties of $\hat{\boldsymbol{\theta}}$. The corresponding expression for $\hat{e}_{\hat{\boldsymbol{\theta}}}$ is

$$V(\hat{e}_{\hat{\boldsymbol{\theta}}}) = V^e_{\boldsymbol{\theta}_0} + \frac{\partial e_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}'}\, \boldsymbol{\Sigma}\, \frac{\partial e_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}.$$

4 Approximations: An example

The discussion so far has been in a multivariate context, i.e. the conditional mean and the conditional covariance function appeared explicitly in the expression for the portfolio returns. We drop that explicitness here and assume that one-period returns are generated by

$$\begin{aligned} y_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \omega + \alpha y^2_{t-1} + \beta h_{t-1} + \gamma \mathbf{1}\left(y_{t-1} < 0\right) y^2_{t-1}, \end{aligned} \tag{12}$$

where ε_t is standard normally distributed and $\mathbf{1}(\cdot)$ is the indicator function. To maintain the portfolio context we can interpret (12) as a process for the cross-sectionally aggregated returns of the assets in the portfolio.[1] Deriving higher moments of temporally aggregated multivariate

[1] Berkowitz and O'Brien (2002) used a similar approach to study the accuracy of the VaR's reported by commercial banks.

GARCH models is technically demanding and to a large extent an unexplored field, though, and we view it as beyond the scope of this particular study (see Hafner, 2003, 2008, for some results).

The conditional variance specification in (12) is the asymmetric GARCH model of Glosten et al. (1993). The term, $\gamma\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$, in (12) extends the basic GARCH$(1,1)$ of Bollerslev (1986) and captures the leverage effect in financial markets, i.e. the asymmetric response of future volatility to positive and negative shocks. This feature has empirically been found highly relevant and several other models to cope with it exist. Wong and So (2003) consider for example the QGARCH model of Sentana (1995) and Engle (1990). The most popular model in empirical work appears, however, to be the GJR-GARCH. In fact, among several different asymmetric GARCH models applied to Japanese stock index data Engle and Ng (1993) found that the best performing parametric specification indeed was the GJR-GARCH.

The implied $VaR_{T,k}^{1-\alpha}$ and $ES_{T,k}^{1-\alpha}$ are given by

$$VaR_{T,k}^{1-\alpha} = -q_{T,k}^{\alpha}\sqrt{h_{T,k}} \tag{13}$$

$$ES_{T,k}^{1-\alpha} = -e_{T,k}^{\alpha}\sqrt{h_{T,k}}. \tag{14}$$

Direct calculation give that the multiple period conditional variance of $Y_{T,k}$ is given by

$$\begin{aligned} h_{T,k} &= \frac{\omega}{1-(\alpha+\beta+\gamma/2)}\left(k - \frac{1-(\alpha+\beta+\gamma/2)^k}{1-(\alpha+\beta+\gamma/2)}\right) \\ &\quad + \frac{1-(\alpha+\beta+\gamma/2)^k}{1-(\alpha+\beta+\gamma/2)} h_{T+1}. \end{aligned}$$

The analytical approximations to $q_{T,k}^{\alpha}$ and $e_{T,k}^{\alpha}$ in (13) and (14) require that we compute theoretical conditional moments of $Y_{T,k}$. We restrict ourselves to the third and the fourth conditional moments and in the Appendix we show how these may be obtained. The corresponding conditional moments of $\varepsilon_{T,k}$ are then

$$s_{T,k} = \frac{E_T(Y_{T,k}^3)}{h_{T,k}^{3/2}},$$

$$k_{T,k} \quad = \quad \frac{E_T(Y_{T,k}^4)}{h_{T,k}^2}.$$

When $\gamma = 0$, the model (12) simplifies to the basic GARCH$(1,1)$ model. Breuer and Jandačka (2007) give expressions for the conditional variance, $h_{T,k}$, and the conditional kurtosis, $k_{T,k}$, of $Y_{T,k}$ under GARCH$(1,1)$ variance.

When $\gamma \neq 0$, there is conditional skewness in $Y_{T,k}$ and the derivation involves non-integer moments $E_T\left(h_{T+k}^{3/2}\right)$ and $E_T\left(h_{T+k}^{5/2}\right)$. Non-integer moments also arise in the context of option pricing in the GARCH framework in Duan, Gauthier, Simonato, and Sasseville (2006), who use Taylor expansions to approximate $E_T\left(h_{T+k}^{1/2}\right)$ and $E_T\left(h_{T+k}^{3/2}\right)$. This is the route taken here as well and the natural starting point for the expansions in our conditional setting is the conditional expectation of the future conditional variance, i.e. $E_T\left(h_{T+k}\right)$. The approximations would then have the form $E_T\left(h_{T+k}^i\right) = a_1 + a_2 E_T\left(h_{T+k}^2\right) + ...$, where $i = 3/2, 5/2$ and the a's are functions of $E_T\left(h_{T+k}\right)$. An important issue is whether higher integer moments of h_{T+k} exist or not for a particular process. Ling and McAleer (2002) derive necessary and sufficient conditions for the unconditional expectation of h_{T+k}^m, m integer, to exist for the family of GARCH$(1,1)$ processes in He and Teräsvirta (1999). The family nests the GJR-GARCH and if $E\left(|\varepsilon|^{2m}\right) < \infty$ the conditions for that particular model are $\omega^m < \infty$ and $E\left[\left(\beta + [\alpha + \gamma \mathbf{1}\left(\varepsilon_{t-1} < 0\right)]\varepsilon_{t-1}^2\right)\right]^m < 1$. The condition for the unconditional variance of y_t to exist is for example $\omega/(1 - \alpha - \beta - \gamma/2) > 0$. Even though the setting here is conditional these conditions can potentially put restrictions on the applicability of our approximation approaches as they require computation of higher moments of y_t. Here, we consider second order expansions.

The approximations based on the Gram-Charlier expansion and the Root-k need no additional comments and are directly obtained by plugging in the expressions for $s_{T,k}$ and $k_{T,k}$. When it comes to choosing a distribution for $\varepsilon_{T,k}$ in the second approach our only requirement is that the first five moments exist for the distribution, and we thus have a large menu to choose from. In the finance literature several distributions have been studied in the context of allowing for conditional skewness and excess kurtosis. Harvey and Siddique (1999) consider a non-central t distribution. Brännäs and Nordman (2003) study the Pearson type IV and the log-generalized gamma. Given that the requirement is satisfied it is difficult to ex ante argue in favor of one distribution over another. A distribution that has gained increasing popularity in the literature (e.g. Jondeau and Rockinger, 2006) is the skewed Student's t distribution of Hansen (1994). Wong and So (2003) propose the distribution in Theodossiou (1998), which is similar to the one in Hansen (1994). They do not pursue the analysis allowing for skewness though and restrict themselves to the symmetric Student's t distribution.

The pdf of a zero mean and unit variance skew-t distributed variable, Z, is

$$g\left(z\right)=\begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^{2}\right)^{-(\nu+1)/2} & \text{if } z<-a/b \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{bz+a}{1+\lambda}\right)^{2}\right)^{-(\nu+1)/2} & \text{if } z\geq -a/b \end{cases},$$

where $2<\nu<\infty$, $-1<\lambda<1$, $a=4\lambda c\left(\nu-2\right)/\left(\nu-1\right)$, $b^{2}=1+3\lambda^{2}-a^{2}$ and $c=\Gamma[(\nu+1)/2]/\left[\sqrt{\pi\left(\nu-2\right)}\Gamma(\nu/2)\right]$. In this particular case the approach consists of matching $s_{T,k}$ and $k_{T,k}$ to the corresponding moments of the skew-t distribution. Jondeau and Rockinger (2003) show that the third and fourth moments of the skew-t distribution are given by

$$\begin{aligned} E\left(Z^{3}\right) &= \left(m_{3}-3a\,m_{2}+2a^{3}\right)/b^{3}, \\ E(Z^{4}) &= \left(m_{4}-4a\,m_{3}+6a^{2}m_{2}-3a^{4}\right)/b^{4}, \end{aligned}$$

where $m_{2}=1+3\lambda^{2}$, $m_{3}=16c\,\lambda(1+\lambda^{2})(\nu-2)^{2}/[(\nu-1)(\nu-3)]$ and $m_{4}=3(\nu-2)(1+10\lambda^{2}+5\lambda^{4})/(\nu-4)$. The third moment is defined for $\nu>3$, while the fourth is defined for $\nu>4$. The implied values on λ and ν are then obtained as the solution in terms of $s_{T,k}$ and $k_{T,k}$ to

$$\begin{aligned} s_{T,k} &= \left(m_{3}-3a\,m_{2}+2a^{3}\right)/b^{3} \\ k_{T,k} &= \left(m_{4}-4a\,m_{3}+6a^{2}m_{2}-3a^{4}\right)/b^{4}. \end{aligned} \tag{15}$$

Except for the symmetric case, i.e. $\lambda = 0$, (when $\nu = (6 - k_{T,k}) / (3 - k_{T,k})$) we were not able to derive λ and ν as nice explicit functions of $s_{T,k}$ and $k_{T,k}$. Obtaining the values then amounts to solving the system numerically.[2] Of course, the valid region for λ and ν also implies a region in the $s_{T,k}$ and $k_{T,k}$ dimension. Jondeau and Rockinger (2003) note that the relation between these regions is bijective when $\nu > 4$ and $|\lambda| < 1$ implying that the solution to (15) is unique.

To compute the VaR and ES we require expressions for $\tilde{q}^{\alpha}_{T,k}$ and $\tilde{e}^{\alpha}_{T,k}$ as inputs to (5) and (6), respectively. Jondeau and Rockinger (2003) show that the αth quantile of the skew-t distribution is given by

$$q_\alpha = \begin{cases} \frac{1}{b}\left[(1-\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha}{1-\lambda}\right) - a\right] & \text{if } \alpha < \frac{1-\lambda}{2} \\ \frac{1}{b}\left[(1+\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha+\lambda}{1+\lambda}\right) - a\right] & \text{if } \alpha \geq \frac{1-\lambda}{2} \end{cases} .$$

In the Appendix we show that, for $\alpha < (1-\lambda)/2$

$$\begin{aligned} E\left(\varepsilon \mid \varepsilon \leq q_\alpha\right) &= -\frac{(1-\lambda)^2}{\alpha b}\left[\frac{\sqrt{v(\nu-2)}}{\nu-1}\left(1+\frac{q_\alpha^{*2}}{\nu}\right) f\left(q_\alpha^*\right)\right. \\ &\quad \left. + \frac{a}{1-\lambda} F\left(q_\alpha^*\right)\right], \end{aligned}$$

where $q_\alpha^* = (bq+a)\sqrt{\nu/(\nu-2)}/(1-\lambda)$ and f and F are the pdf and the cdf of the Student's t distribution.

Quantifying the uncertainty of the predictors follows from Section 3.

[2] In the simulation study in Section 5 we employed the following solver

$$\begin{aligned} (\lambda, \nu)' &= \underset{-1<\lambda<1, \nu>4}{\arg\min} \left[s_{T,k} - \left(m_3 - 3a\, m_2 + 2a^3\right)/b^3\right]^2 \\ &\quad + \left[k_{T,k} - \left(m_4 - 4a\, m_3 + 6a^2 m_2 - 3a^4\right)/b^4\right]^2 . \end{aligned}$$

Issues with using a solver of this type are discussed in Press, Teukolsky, Vetterling, and Flannery (2007, ch. 9). However, it performed satisfactory in our application with function values close to zero. We also compared it to the Newton-Raphson algorithm in Press et al. (2007, ch. 9, p. 475) and almost identical values were obtained. The latter was highly sensitive to the starting values, though.

5 Simulation study

The discussion regarding the approximative predictors has so far been theoretical, but what is of obvious practical interest is their properties in finite sample. We address this question by means of quite detailed Monte Carlo simulations based on the model in (12). The study was carried out using the RATS 6.30 package. To estimate the GJR-GARCH models we employed the built-in GARCH procedure with the BFGS-algorithm, but as the variance-covariance estimator we used $T^{-2}(\partial \ln L_T(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}$ $\partial \ln L_T(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}')$.

When it comes to designing the experiment we note for the variance specification that the degree of persistence and asymmetry are of particular interest. In a related study Christoffersen and Gonçalves (2005) simulate the GARCH(1, 1)-model with $\omega = (1 - \alpha - \beta)20^2/252$, $\alpha = 0.1$ and persistence parameter $\beta = 0.4, 0.8$ and 0.89. Here, the additional parameter γ introduces asymmetry and we consider three degrees: $(\alpha, \gamma) = (0.1, 0)$, $(0.05, 0.1)$ and $(0, 0.2)$. The unconditional variance is thus the same throughout. For estimation we use samples of sizes 500 and 1000, which are realistic sample sizes corresponding to approximately 2 and 4 years of daily trading data. For the simulation based predictor we use $R = 100\ 000$ to isolate the effect of the estimation error in $\hat{\boldsymbol{\theta}}$. The results are based on $N = 1000$ replications. Note however, that we discard without replacement the cases when the ML estimator did not converge to a valid point or when an approximation failed for some reason. Table A1 in the Appendix gives the proportions of cases when this happened. The remaining design parameters are the confidence level and the horizon. Increasing the confidence level means that we make predictions further out in the left tail, which intuitively increases the uncertainty. Predicting further into the future is also associated with greater uncertainty, which should be reflected in the performance of the predictors. We set the confidence level to either 95% or 99% and consider $k = 5$ and 10. In Table 1 we give bias, mean square errors (MSE) and estimated asymptotic variances (EAV) for the case $\beta = 0.8$ and $(\alpha, \gamma) = (0.05, 0.1)$. The tables for the other parameter combinations are given in the Appendix.

We make no distinction between VaR and ES in the discussion as the results are qualitatively similar. Considering first the bias we see that it is largest and negative for the Root-k approach. The bias for the G-C approach is positive for all cases and surprisingly large for the higher confidence level. Overall, it is the smallest for the W-S and the simulation based approaches. With some exceptions, the bias gets more pronounced when increasing the confidence level and the horizon, and it decreases when increasing the sample size. Turning to the accuracy in terms of MSE we again have a rather clear ranking with the Root-k approach being the worst and the W-S and the simulation based approach tied in first place. Without exceptions, the qualitative effects of the design variables are the same as for the bias. Of interest for the computation of, e.g., prediction intervals is how well the delta method approximates the finite sample variance of the predictors. To scrutinize on this we may compare the MSE to the average of the corresponding estimated asymptotic variances.[3] The delta method appears to perform quite satisfactorily for the G-C, the W-S and the simulation based approaches. Regarding the Root-k approach it is difficult to draw any conclusions due to the often large bias.

In a smaller scale experiment we examined the robustness of the results for data generated according to a GJR-GARCH process ($\alpha = .05$, $\beta = 0.8$ and $\gamma = 0.1$) and with skew–t innovations ($\lambda = -0.2$ and $\nu = 8$). We computed the predictors for a confidence level of 95% and $k = 5$ based on $T = 1000$ both with a correct distributional assumption and with an incorrect assumption of normality (cf. QMLE). Regarding the former the predictors need to be adapted to the skew-t distribution and the corresponding derivations may be found in the Appendix.

[3] Here we rely on a central limit theorem argument. Let $\{Z_n\}_{n=1}^N$ be independently distributed variables with zero means and variances σ_n^2. Then $\sum_{n=1}^N Z_n^2/N$ is a consistent estimator of $\bar{\sigma}^2 = \sum_{n=1}^N \sigma_n^2/N$. Hence, when the bias is small the average of the estimated asymptotic variances should be close to the MSE.

Table 1: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim \text{nid}(0,1)$ and $h_t = \omega + 0.05y_{t-1}^2 + 0.8h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance. The averages are for the true values.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Average	4.7041			6.7147			6.2311			8.9773		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
					$T=500$							
Root-k	-0.1109	0.1524	0.1243	-0.2200	0.4417	0.2486	-0.4707	0.4247	0.1955	-0.8322	1.2222	0.3910
G-C	0.0357	0.1401	0.1404	0.0675	0.3247	0.3081	0.2340	0.5080	0.5319	0.5437	1.8331	1.8692
W-S	0.0280	0.1364	0.1329	0.0446	0.3062	0.2810	-0.0058	0.2776	0.2818	-0.0068	0.6915	0.6743
Sim	-0.0115	0.1301	0.1484	-0.0203	0.2876	0.3447	-0.0195	0.2753	0.2869	-0.0429	0.6780	0.6704
					$T=1000$							
Root-k	-0.0995	0.0734	0.0572	-0.2028	0.2723	0.1145	-0.4567	0.2931	0.0900	-0.8112	0.9393	0.1801
G-C	0.0456	0.0605	0.0652	0.0873	0.1454	0.1446	0.2231	0.2403	0.2238	0.5176	0.9450	0.7638
W-S	0.0387	0.0586	0.0621	0.0669	0.1348	0.1326	0.0091	0.1199	0.1304	0.0243	0.3057	0.3154
Sim	-0.0010	0.0545	0.0673	-0.0007	0.1238	0.1577	-0.0013	0.1179	0.1339	-0.0063	0.2994	0.3164

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Average	7.1562			10.3360			8.6290			12.5751		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.6596	0.6892	0.2486	-1.1496	1.9641	0.4973	-1.1858	1.7309	0.3264	-2.0496	4.9457	0.6527
G-C	0.1247	0.5109	0.5478	0.2657	1.5148	1.5921	0.3088	0.9728	0.9175	0.3404	2.1102	2.1023
W-S	-0.0211	0.3923	0.3982	-0.0259	0.9982	0.9835	-0.0896	0.6902	0.7208	-0.1418	1.9228	1.9354
Sim	-0.0231	0.3947	0.4470	-0.0555	0.9890	1.1618	-0.0303	0.7290	0.7875	-0.0808	1.9926	2.0277
						$T=1000$						
Root-k	-0.6438	0.5185	0.1145	-1.1261	1.5978	0.2290	-1.1681	1.4975	0.1503	-2.0237	4.4416	0.3006
G-C	0.1358	0.2279	0.2434	0.2829	0.6916	0.6898	0.3652	0.5425	0.4611	0.5443	1.2579	1.0328
W-S	-0.0031	0.1706	0.1839	0.0116	0.4435	0.4598	-0.0688	0.3047	0.3285	-0.0998	0.8602	0.9024
Sim	-0.0008	0.1700	0.2055	-0.0084	0.4368	0.5428	0.0010	0.3173	0.3656	-0.0147	0.8926	0.9634

Table 2: Diebold-Mariano t-tests. Positive values are in favor of the method in the second row. The loss function is the squared prediction error and the statistics were computed from the regression of the pooled differences on a constant using Eicker-White standard errors. The average differences are given in parentheses.

		Root-k	Root-k	Root-k	G-C	G-C	W-S
	vs.	G-C	W-S	Sim	W-S	Sim	Sim
VaR		48.441	72.704	74.094	30.599	31.051	8.980
		(0.351)	(0.467)	(0.471)	(0.117)	(0.121)	(0.005)
ES		50.354	107.547	107.019	44.180	43.808	-4.231
		(0.853)	(1.396)	(1.390)	(0.542)	(0.536)	(-0.004)

The results are given in Table A10 of the Appendix and, with some exceptions, they are qualitatively similar to the ones above when the model is correctly specified. However, it is noteworthy that the delta method appears to work poorly in many cases. Also, the bias and the MSE for the G-C approach in case of ES prediction is very high. Under an incorrect normality assumption the bias is negative in all cases and the ranking is different.

An important question we wish to answer is that of which method is the best. For this we use a Diebold-Mariano type of test (Diebold and Mariano, 1995). They show that the predictive superiority of one predictor over another can be tested by means of a simple t-test of the standardized difference between the loss functions. Here, the loss function is the squared prediction error and the test statistic was computed as the t-statistic in the regression of the pooled differences on a constant. To take care of heteroskedasticity we used Eicker-White standard errors. In Table 2 we give results for all pairwise tests.

Among the analytical approaches the one based on the skew-t distribution is judged the best. In fact, it also fares better than the simulation based in case of ES prediction. For VaR, all analytical approaches are rejected in favor of the simulation based. The actual differences between the simulation based and the one based on the skew-t is small, though. In a practical situation one would thus supposedly prefer the latter thanks to its advantage in computing time. For example, consider the task of computing VaR and ES for confidence levels 95% and 99% and horizons of 5 and 10 periods given that parameter estimates have been obtained. Along with standard errors it takes approximately 25 seconds on a 1.83 GHz Intel Centrino Duo processor employing the simulation based approach, while the other approaches compute the quantities within the blink of an eye.

Of further practical interest is how the prediction accuracy varies with the design variables, i.e. the sample size, confidence level, horizon and model parameters. For this we ran the dummy variable regressions

$$\begin{aligned} \ln(\widehat{VaR}_{T,k}^{1-\alpha} - VaR_{T,k}^{1-\alpha})^2 &= \boldsymbol{\beta}_v' \mathbf{d} + \xi_v, \\ \ln(\widehat{ES}_{T,k}^{1-\alpha} - ES_{T,k}^{1-\alpha})^2 &= \boldsymbol{\beta}_e' \mathbf{d} + \xi_e, \end{aligned}$$

where $\mathbf{d}$ is a vector of dummy variables indicating value on design variable and ξ_v and ξ_e are the error terms. We again used Eicker-White standard errors. The base case is taken to be $T = 500$, $\alpha = .05$, $k = 5$, $\beta = .8$ and $\gamma = .1$. We ran one regression for each method and the results are given in Table 3.

The results were uniform across the methods and qualitatively the same for VaR and ES. Not surprisingly, doubling the sample size significantly increased the accuracy. The predictions at confidence level 99% were significantly less accurate than the ones at the 95% level. Increasing the horizon from 5 to 10 periods significantly decreased the accuracy. Both when increasing and reducing the persistence in the conditional variance the accuracy is significantly enhanced compared to the base case. Regarding the effects of asymmetry we note that predictions for the no asymmetry case (i.e. standard GARCH) are significantly more accurate than the ones for the base case. The opposite is true for the case when only negative shocks affects the future variance.

Table 3: Effects of design variables on the accuracy of the predictors. The numbers are the values on t-tests of zero coefficents in dummy variable regressions, where the base case is $T = 500$, $\alpha = 0.05$, $k = 5$, $\beta = 0.8$ and $\gamma = 0.1$.

	Root-k		G-C		W-S		Sim	
Dummy	VaR	ES	VaR	ES	VaR	ES	VaR	ES
$T = 1000$	-12.438	-6.576	-33.782	-24.312	-37.497	-41.051	-39.098	-40.875
$\alpha = .01$	147.422	148.049	79.504	34.673	64.808	56.867	66.674	59.286
$k = 10$	79.897	90.368	52.018	54.081	47.605	50.566	47.553	50.450
$\beta = .4$	-17.208	-47.475	-28.518	-31.511	-22.629	-22.056	-24.940	-22.676
$\beta = .89$	-19.582	-11.402	-16.423	-12.319	-16.621	-19.465	-15.831	-19.242
$\gamma = .0$	-55.443	-88.258	-23.289	-38.378	-19.264	-26.797	-16.953	-25.971
$\gamma = .2$	45.713	67.273	11.665	25.276	4.441	7.945	2.382	5.776

6 Empirical illustration

In this section we provide a small illustration of the above approximation approaches, where the object of interest is the five day VaR of the S&P 500 index. Eight years of daily data were downloaded from DataStream and the sample covers October 31, 2000 to October 31, 2008, for a total of 2089 daily observations on the index.

Returns were calculated as $y_t = 100 \times \log(I_t/I_{t-1})$, where I_t is the value of the index at t. We assume that daily returns are generated by a GJR-GARCH process with a constant mean and standard normally distributed shocks. In the estimation of the model as is we often obtained a negative coefficient on the squared residual. This causes problems for the simulation based predictor, since the conditional variance may become negative in the out of sample simulations. To force positive variances we adopted the following version

$$h_t = \omega + \exp(\alpha)u_{t-1}^2 + \beta h_{t-1} + \gamma \mathbf{1}\left(u_{t-1} < 0\right) u_{t-1}^2,$$

where u_t is the one-period ahead prediction error. Regarding the computation of VaR a comment is in place. Recall the decomposition $Y_{T,k} = \mu_{T,k} + \varepsilon_{T,k}\sqrt{h_{T,k}}$. As inputs to the G-C and W-S approximations we require the conditional skewness and kurtosis of $\varepsilon_{T,k}$. Those were derived in the Appendix under a zero conditional mean of $Y_{T,k}$. Here, we use the same derivation but replace $Y_{T,k}$ with $Y_{T,k} - \mu_{T,k}$, where $\mu_{T,k} = k\mu$. Note also that the uncertainty in μ should be recognized in the computation of the variances of VaR.

Based on a rolling prediction scheme we obtained VaR predictions at the confidence level 95% and in estimation we considered samples of size 500 observations. We discarded cases when the computation of the predictors failed for some reason and obtained 1522 predictions. Robust standard errors of the sandwich form were employed throughout. The final successfully estimated model (October 3, 2008) on the implied conventional form is given below along with some diagnostics.

$$
\begin{aligned}
y_t &= \underset{(-0.52)}{-0.018} + u_t, \\
h_t &= \underset{(1.95)}{0.014} + \underset{(0.10)}{1 \times 10^{-5}} u_{t-1}^2 + \underset{(36.88)}{0.891} h_{t-1} + \underset{(4.99)}{0.211} \mathbf{1}(u_{t-1} < 0) u_{t-1}^2, \\
L &= -662.24,\ LB_{10} = 8.22,\ LB_{10}^2 = 6.95,\ JB = 0.29,
\end{aligned}
$$

where t-statistics are given in parentheses, L is the value of the log-likelihood function, LB_{10} and LB_{10}^2 give the values of the test-statistics in the Ljung-Box test of no autocorrelation up to lag 10 in standardized residuals and squared standardized residuals, respectively, and JB is the value of the test-statistic in the Jarque-Bera normality-test. The conditional variance is highly persistent and the asymmetric effect of past shocks is considerable. Noteworthy is also that there is no remaining ARCH-effect in the standardized residuals and that normality is not rejected.

When it comes to assessing the performance of the VaR predictors we follow the likelihood ratio framework of Christoffersen (1998). Let P denote the number of VaR predictions and let H_t, $t = 1, ..., P$, denote the hit sequence, i.e. $H_t = 1$ if the actual return exceeds the predicted

VaR and is 0 otherwise. For a good VaR predictor the unconditional exceedence rate, $\hat{\alpha} = \sum H_t/P$, should be close to α. This can be tested by the statistic $LR_{unc} = -2\ln[(1-\alpha)^{P-H}\alpha^H] + 2\ln[(1-\hat{\alpha})^{P-H}\hat{\alpha}^H]$. Christoffersen (1998) notes that the hit sequence should not only sum up to αP, but also be an iid Bernoulli sequence with parameter α. As a test of independence he proposes the test statistic $LR_{ind} = -2\ln[(1-\hat{\alpha})^{P_{00}+P_{10}}\hat{\alpha}^{P_{01}+P_{11}}] + 2\ln[(1-\hat{\pi}_0)^{P_{00}}\hat{\pi}_0^{P_{01}}(1-\hat{\pi}_1)^{P_{10}}\hat{\pi}_1^{P_{11}}]$, where P_{ij} is defined as the number of periods in which state j occurred in one period, while state i occurred the previous period and π_i is the probability of a hit conditional on state i the previous day.[4] He proposes $LR_{cc} = LR_{unc} + LR_{ind}$ as a statistic for the joint test of correct conditional coverage. Asymptotically LR_{unc} and LR_{ind} are $\chi^2(1)$-distributed, while LR_{cc} is $\chi^2(2)$-distributed. Our multiple period context may give rise to serial dependence in the raw hit sequence. To cope with the problem we use Bonferroni subsamples (Dowd, 2007). Thus, the raw hit sequence is split up into five hit sequences and the statistics are computed for each sequence. We reject an overall test at significance level λ if the test is rejected for any of the subsamples using level $\lambda/5$.

The unconditional exceedence rates for the predictors are 5.532%, 5.506%, 5.506% and 5.512% for the Root-k, G-C, W-S and simulated based approach, respectively. In Figure 1 we display VaR's and standard errors for the turbulent period September 5, 2008 to October 3, 2008.

To digress further on the performance of the predictors we present in Table 4 the results of the backtesting of the VaR predictors. The approaches perform very similarly and no results are significant at conventional levels. Note that the performances of all predictors are quite weak when the prediction origin is a Monday (too many hits) or a Thursday (too few hits).

[4] When $\hat{\pi}_1 = 0$ we used $LR_{ind} = -2\ln[(1-\hat{\alpha})^{P_{00}+P_{10}}\hat{\alpha}^{P01+P11}] + 2\ln[(1-\hat{\pi}_0)^{P_{00}}\hat{\pi}_0^{P_{01}}$ (cf. Christoffersen and Pelletier, 2004).

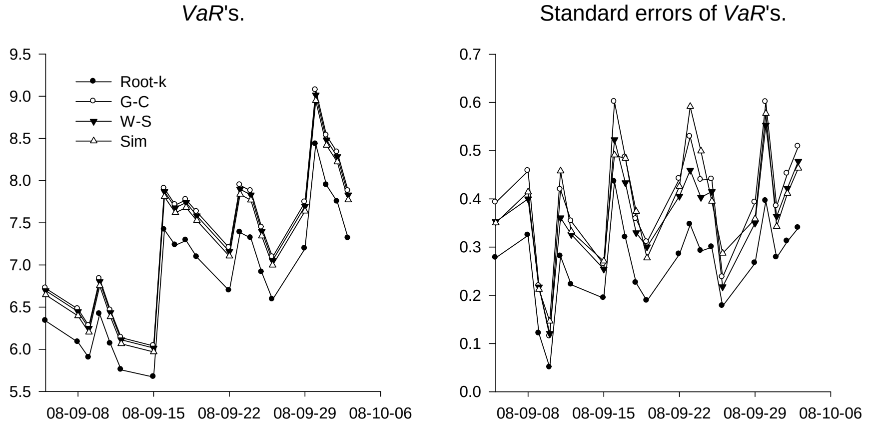

Figure 1: VaR's for 100× log-returns for the S&P 500 index.

Table 4: Backtesting of the VaR predictors. The top row indicate day of the week of the prediction origin.

	Root-k					G-C				
	M	T	W	T	F	M	T	W	T	F
$\hat{\alpha}$	0.0689	0.0498	0.0525	0.0391	0.0559	0.0656	0.0465	0.0492	0.0391	0.0526
LR_{unc}	2.0525	0.0002	0.0382	0.8293	0.2165	1.4244	0.0789	0.0043	0.8293	0.0436
LR_{ind}	0.0540	0.1275	0.0640	0.8782	1.6185	0.0311	1.2665	1.4333	0.8782	1.4223
LR_{cc}	2.1064	0.1277	0.1022	1.7075	1.8350	1.4555	1.3453	1.4376	1.7075	1.4659

	W-S					Sim				
	M	T	W	T	F	M	T	W	T	F
$\hat{\alpha}$	0.0656	0.0465	0.0492	0.0391	0.0526	0.0656	0.0465	0.0492	0.0391	0.0559
LR_{unc}	1.4244	0.0789	0.0043	0.8293	0.0436	1.4244	0.0789	0.0043	0.8293	0.2165
LR_{ind}	0.0311	1.2665	1.4333	0.8782	1.4223	0.0311	1.2665	1.4333	0.8782	1.6185
LR_{cc}	1.4555	1.3453	1.4376	1.7075	1.4659	1.4555	1.3453	1.4376	1.7075	1.8350

7 Conclusions

In this paper we studied four methods to approximate VaR and ES for multiple period returns. We also viewed the uncertainty arising from the estimation error important and we discussed how to employ the delta method to quantify this uncertainty. Based on the result of a simulation experiment we conclude that among the approaches studied the one based on assuming a skew-t distribution for the multiple period returns and that based on simulations were the best. The predictors based on the Root-k and the Gram-Charlier showed positive and negative bias, respectively. Except for the Root-k approach we found that the uncertainty due to the estimation error can be quite accurately estimated employing the delta method.

In an empirical illustration we computed 5 day VaR's for the S&P 500 index using the approximative predictors. In terms of exceedence rates all approaches performed similarly and we could not reject any of them at conventional significance levels.

Appendix

Conditional moments of $Y_{T,k}$ with GJR-GARCH conditional variance

We consider here the case $E_T(y_{T+i}) = 0$ and when deriving the conditional moments it is helpful to use the decomposition $Y_{T,k} = \sum_{i=1}^{k-1} y_{t+i} + y_{t+k}$. Let $s = E(\varepsilon^3_{T+i})$ and $\kappa = E(\varepsilon^4_{T+i})$. For notational convenience we let $\varepsilon = \varepsilon_{T+i}$. Obtaining the moments then amounts to solving the system

$$
\begin{aligned}
E_T\left[\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2\right] &= E_T\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2 + E_T\left(h_{T+k-1}\right) && \text{(A1)}\\
E_T\left[\left(\sum_{i=1}^{k} y_{T+i}\right)^3\right] &= sE_T(h^{3/2}_{T+k}) + E_T\left(\sum_{i=1}^{k-1} y_{T+i}\right)^3 \\
&\quad +3E_T\left(h_{T+k}\sum_{i=1}^{k-1} y_{T+i}\right) && \text{(A2)}\\
E_T\left[\left(\sum_{i=1}^{k} y_{T+i}\right)^4\right] &= \kappa E_T h^2_{T+k} + E_T\left(\sum_{i=1}^{k-1} y_{T+i}\right)^4 \\
&\quad +4sE_T\left(h^{3/2}_{T+k}\sum_{i=1}^{k-1} y_{T+i}\right) \\
&\quad +6E_T h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2 && \text{(A3)}\\
E_T\left(h_{T+k}\right) &= \omega + \delta E_T h_{T+k-1} && \text{(A4)}\\
E_T\left(h^{3/2}_{T+k}\right) &\approx \frac{5}{8}\left(E_T h_{T+k}\right)^{3/2} \\
&\quad + \frac{3}{8\sqrt{E_T h_{T+k}}} E_T h^2_{T+k} && \text{(A5)}\\
E_T\left(h^2_{T+k}\right) &= \omega^2 + 2\omega\delta E_T\left(h_{T+k-1}\right) \\
&\quad +\lambda E_T h^2_{T+k-1} && \text{(A6)}
\end{aligned}
$$

$$
\begin{aligned}
E_T\left(h_{T+k}^{5/2}\right) &\approx -\frac{7}{8}(E_T h_{T+k})^{5/2} \\
&\quad +\frac{15}{8}\sqrt{E_T h_{T+k}}E_T h_{T+k}^2 &\text{(A7)} \\
E_T\left(h_{t+k}\sum_{i=1}^{k-1} y_{t+i}\right) &= [\alpha s+\gamma E(\mathbf{1}(\varepsilon<0)\varepsilon^3)]E_T h_{T+k-1}^{3/2} \\
&\quad +\delta E_T\left(h_{k+t-1}\sum_{i=1}^{k-2} y_{i+t}\right) &\text{(A8)} \\
E_T\left[h_{T+k}^{3/2}\left(\sum_{i=1}^{k-1} y_{T+i}\right)\right] &\approx \frac{3}{4}(E_T h_{T+k})^{1/2}E_T\left[h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)\right] \\
&\quad +\frac{3}{8\sqrt{E_T h_{T+k}}}E_T\left(h_{T+k}^2\sum_{i=1}^{k-1} y_{T+i}\right) &\text{(A9)} \\
E_T\left(h_{T+k}^2\sum_{i=1}^{k-1} y_{T+i}\right) &= 2\omega\delta E_T\left(h_{T+k-1}\sum_{i=1}^{k-2} y_{T+i}\right) \\
&\quad +\lambda E_T\left(h_{T+k-1}^2\sum_{i=1}^{k-2} y_{T+i}\right) \\
&\quad +\pi E_T\left(h_{T+k-1}^{5/2}\right) \\
&\quad +[2\alpha\omega s+2\gamma\omega E(\mathbf{1}(\varepsilon<0)\varepsilon^3)] \\
&\quad \times E_T\left(h_{T+k-1}^{3/2}\right) &\text{(A10)} \\
E_T\left[h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2\right] &= \omega E_T\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2 \\
&\quad +\delta E_T\left[h_{T+k-1}\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2\right] \\
&\quad +[2\alpha s+2\gamma E(\mathbf{1}(\varepsilon<0)\varepsilon^3)] \\
&\quad \times E_T\left[h_{T+k-1}^{3/2}\left(\sum_{i=1}^{k-2} y_{T+i}\right)\right] \\
&\quad +\omega E_T h_{T+k-1}+\mu E_T\left(h_{T+k-1}^2\right), &\text{(A11)}
\end{aligned}
$$

where $\mathbf{1}(\cdot)$ is the indicator function, $\delta = \alpha + \beta + \gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^2)$, $\lambda = \beta^2 + \kappa\alpha^2 + 2\alpha\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^4) + 2\alpha\beta + 2\beta\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^2) + \gamma^2 E(\mathbf{1}(\varepsilon < 0)\varepsilon^4)$, $\mu = \kappa\alpha + \beta + \gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^4)$ and $\pi = \alpha^2 E(\varepsilon^5) + 2\alpha\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^5) + 2\alpha\beta s + 2\beta\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^3) + \gamma^2 E(\mathbf{1}(\varepsilon < 0)\varepsilon^5)$.

When ε is Gaussian we have $s = E(\varepsilon^5) = 0$ and $E(\varepsilon^4) = 3$. Also, it is straightforward to show that $E(\mathbf{1}(\varepsilon < 0)\varepsilon) = -\phi(0)$. For integer $r > 1$ it holds that $\int_{-\infty}^0 z^r\phi(z)dz = (r-1)\int_{-\infty}^0 z^{r-2}\phi(z)dz$. We have $E(\mathbf{1}(\varepsilon < 0)\varepsilon^2) = 1/2$, $E(\mathbf{1}(\varepsilon < 0)\varepsilon^3) = -2\phi(0)$, $E(\mathbf{1}(\varepsilon < 0)\varepsilon^4) = 3/2$ and $E(\mathbf{1}(\varepsilon < 0)\varepsilon^5) = -8\phi(0)$.

Properties of the skew-t distribution

We take a, b, c, m_2, m_3 and m_4 as they are given in the text. Jondeau and Rockinger (2003) give the αth quantile of the skew-t distribution as

$$q_\alpha = \begin{cases} \frac{1}{b}\left[(1-\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha}{1-\lambda}\right) - a\right] & \text{if } \alpha < \frac{1-\lambda}{2} \\ \frac{1}{b}\left[(1+\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha+\lambda}{1+\lambda}\right) - a\right] & \text{if } \alpha \geq \frac{1-\lambda}{2} \end{cases},$$

where $F^{-1}(\cdot)$ is the inverse of the cdf of the Student's t distribution with ν degrees of freedom.

To solve the system (A1) - (A11) we require some integer moments. We first derive the censored ones for the standardized Student's t distribution. Let $\mu_m^q = \int_{-\infty}^q x^m t(x)\,dx$, where $t(x) = c\,[1+x^2/(\nu-2)]^{-(\nu+1)/2}$. μ_0^q is obvious and adapting a result in Andreev and Kanto (2005) gives

$$\mu_1^q = -\frac{\nu-2}{\nu-1}\left(1+\frac{q^2}{\nu-2}\right)t(q).$$

By integration by parts we have for $m > 1$

$$\begin{aligned} \mu_m^q &= \int_{-\infty}^q x^m c\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2} dx \\ &= \left\{-x^{m-1}\frac{\nu-2}{\nu-1}\left[\left(1+\frac{x^2}{\nu-2}\right)t(x)\right]\right\}_{-\infty}^q \\ &\quad +(m-1)\frac{\nu-2}{\nu-1}\int_{-\infty}^q x^{m-2}\,c\left(1+\frac{x^2}{\nu-2}\right)t(x)dx \end{aligned}$$

$$= \quad q^{m-1}\mu_1^q + (m-1)\frac{\nu-2}{\nu-1}\left(\mu_{m-2}^q + \frac{\mu_m^q}{\nu-2}\right).$$

Then

$$\mu_m^q = \frac{\nu-1}{\nu-m}\left(q^{m-1}\mu_1^q + (m-1)\frac{\nu-2}{\nu-1}\mu_{m-2}^q\right).$$

Now, for the skew-t distributed variable Z we have for $q < -a/b$

$$\begin{aligned}
E(\mathbf{1}(Z \le q)Z^m) &= \int_{-\infty}^{q} z^m bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} dz \\
&= \frac{\lambda_*}{b^m}\int_{-\infty}^{q_*}[\lambda_* y - a)]^m t(y)dy \\
&= \frac{\lambda_*}{b^m}\int_{-\infty}^{q_*}\sum_{i=0}^{m}\binom{m}{i}\lambda_*^{m-i}(-a)^i y^{m-i} t(y)dy \\
&= \frac{\lambda_*}{b^m}\sum_{i=0}^{m}\binom{m}{i}\lambda_*^{m-i}(-a)^i\mu_{m-i}^{q*},
\end{aligned}$$

where we use a change of variable $y = (bz+a)/(1-\lambda)$ in the first step, and where $\lambda_* = 1-\lambda$ and $q_* = (bq+a)/(1-\lambda)$. We obtain

$$\begin{aligned}
E(\mathbf{1}(Z \le q)Z) &= \frac{\lambda_*}{b}(\lambda_*\mu_1^{q*} - a\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^2) &= \frac{\lambda_*}{b^2}(\lambda_*^2\mu_2^{q*} - 2a\lambda_*\mu_1^{q*} + a^2\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^3) &= \frac{\lambda_*}{b^3}(\lambda_*^3\mu_3^{q*} - 3a\lambda_*^2\mu_2^{q*} + 3a^2\lambda_*\mu_1^{q*} \\
&\quad -a^3\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^4) &= \frac{\lambda_*}{b^4}(\lambda_*^4\mu_4^{q*} - 4a\lambda_*^3\mu_3^{q*} + 6a^2\lambda_*^2\mu_2^{q*} \\
&\quad -4a^3\lambda_*\mu_1^{q*} + a^4\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^5) &= \frac{\lambda_*}{b^5}(\lambda_*^5\mu_5^{q*} - 5a\lambda_*^4\mu_4^{q*} + 10a^2\lambda_*^3\mu_3^{q*} \\
&\quad -10a^3\lambda_*^2\mu_2^{q*} + 5a^4\lambda_*\mu_1^{q*} - a^5\mu_0^{q*}.
\end{aligned}$$

Note that in the computation of the ES we use $E\left(Z \mid Z \le q_a\right) = E(\mathbf{1}(Z \le q_a)Z)/\alpha$.

For $E\left(Z^5\right)$ we build on Jondeau and Rockinger (2003), who rely on the result of Gradshteyn and Ryzhik (1994):

$$\int_0^\infty x^{\mu-1}(p+qx^\nu)^{-(n+1)}dx = \frac{1}{\nu p^{n+1}}\left(\frac{p}{q}\right)^{\mu/\nu} \times \frac{\Gamma(\mu/\nu)\Gamma[1+n-(\mu/\nu)]}{\Gamma(1+n)}, \tag{A12}$$

where $0 < \mu/\nu < n+1$, $p \neq 0$, $q \neq 0$, $\Gamma(\cdot)$ is the gamma function with $\Gamma(x) = (x-1)\Gamma(x-1)$ and $\Gamma(1/2) = \sqrt{\pi}$.

Consider the variable $Y = Za + b$ with density

$$h\left(y\right) = \begin{cases} c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{if } y \leq 0, \\ c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{if } y > 0. \end{cases}$$

We have

$$\begin{aligned} E(Y^5) &= m_5 = \int_{-\infty}^0 y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} dy \\ &\quad + \int_0^\infty y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} dy \\ &= I_l + I_r, \end{aligned}$$

and on using (A12) we get

$$\begin{aligned} I_l &= \int_{-\infty}^0 y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{(1-\lambda)}\right)^2\right)^{-(\nu+1)/2} dy \\ &= c(1-\lambda)^6 \int_{-\infty}^0 x^5\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2} dx \\ &= -c(1-\lambda)^6(\nu-2)^3\frac{\Gamma[(\nu-5)/2]}{\Gamma[(\nu+1)/2]} \\ &= -8c(1-\lambda)^6\frac{(\nu-2)^3}{(\nu-1)(\nu-3)(\nu-5)}, \end{aligned}$$

where we use the change of variable $x = y/(1-\lambda)$ in the first step. Similarly, $I_r = 8c(1+\lambda)^6(\nu-2)^3/[(\nu-1)(\nu-3)(\nu-5)]$ and $m_5 = 8c(\nu-2)^3[(1+\lambda)^6-(1-\lambda)^6]/[(\nu-1)(\nu-3)(\nu-5)]$. We then have

$$E\left(Z^5\right) = \frac{E(Y-a)^5}{b^5} = \frac{m_5 + 4a^5 - 5am_4 - 10a^3m_2 + 10a^2m_3}{b^5}.$$

Table A1: Proportions of cases when the ML estimator did not converge to a valid point or when the indicated approximation failed. The reported numbers are maxima taken over the considered confidence levels and horizons.

	$\beta = .4$				$\beta = .8$				$\beta = .89$			
	Root-k	G-C	W-S	Sim	Root-k	G-C	W-S	Sim	Root-k	G-C	W-S	Sim
$T = 500$												
$\gamma = 0$	0.016	0.016	0.016	0.016	0.001	0.001	0.001	0.001	0.074	0.074	0.074	0.074
$\gamma = 0.1$	0.012	0.011	0.011	0.011	0.008	0.008	0.012	0.008	0.111	0.111	0.111	0.111
$\gamma = 0.2$	0.017	0.017	0.048	0.017	0.013	0.013	0.018	0.013	0.176	0.177	0.176	0.176
$T = 1000$												
$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.039	0.039	0.039
$\gamma = 0.1$	0.003	0.003	0.003	0.003	0.000	0.000	0.000	0.000	0.056	0.056	0.056	0.056
$\gamma = 0.2$	0.008	0.007	0.009	0.007	0.001	0.006	0.002	0.012	0.085	0.090	0.086	0.100

Table A2: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.1 y_{t-1}^2 + 0.4 h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
$T=500$												
Root-k	-0.0011	0.1199	0.0744	-0.0178	0.3038	0.1487	-0.1385	0.2037	0.1169	-0.1560	0.4974	0.2339
G-C	-0.0016	0.0420	0.0438	-0.0050	0.0747	0.0784	0.0091	0.0924	0.0948	0.0085	0.1507	0.1616
W-S	-0.0027	0.0416	0.0443	-0.0052	0.0744	0.0794	-0.0062	0.0830	0.0879	-0.0015	0.1418	0.1541
Sim	-0.0069	0.0437	0.1017	-0.0108	0.0774	0.1531	-0.0129	0.0880	0.0996	-0.0104	0.1481	0.1756
$T=1000$												
Root-k	0.0035	0.0727	0.0362	-0.0114	0.2031	0.0723	-0.1330	0.1282	0.0569	-0.1484	0.3375	0.1137
G-C	0.0055	0.0214	0.0213	0.0050	0.0363	0.0372	0.0146	0.0433	0.0445	0.0151	0.0695	0.0740
W-S	0.0027	0.0213	0.0215	0.0027	0.0363	0.0375	-0.0029	0.0407	0.0419	0.0022	0.0673	0.0715
Sim	0.0017	0.0219	0.0347	0.0018	0.0379	0.0672	-0.0055	0.0416	0.0437	-0.0012	0.0698	0.0741

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2067	0.2756	0.1487	-0.2245	0.6516	0.2975	-0.4261	0.4825	0.1952	-0.4414	0.9796	0.3904
G-C	0.0288	0.1430	0.1469	0.0274	0.2194	0.2392	0.1364	0.3947	0.3974	0.1192	0.5137	0.5771
W-S	-0.0072	0.1154	0.1223	0.0025	0.1943	0.2127	-0.0129	0.2127	0.2277	0.0068	0.3405	0.3784
Sim	-0.0154	0.1221	0.1798	-0.0103	0.2034	0.3564	-0.0215	0.2295	0.2622	-0.0038	0.3615	0.4422
						$T=1000$						
Root-k	-0.2006	0.1793	0.0723	-0.2161	0.4476	0.1446	-0.4192	0.3543	0.0949	-0.4319	0.7102	0.1899
G-C	0.0316	0.0647	0.0668	0.0305	0.0986	0.1067	0.1236	0.1722	0.1737	0.1010	0.2168	0.2374
W-S	-0.0044	0.0565	0.0582	0.0046	0.0916	0.0983	-0.0153	0.1025	0.1065	-0.0001	0.1592	0.1715
Sim	-0.0081	0.0579	0.0804	-0.0024	0.0943	0.1702	-0.0191	0.1063	0.1123	-0.0057	0.1671	0.1793

Table A3: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.05y_{t-1}^2 + 0.4h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
$T=500$												
Root-k	-0.0838	0.1872	0.1097	-0.1350	0.4496	0.2194	-0.3367	0.3904	0.1725	-0.4266	0.8522	0.3451
G-C	0.0176	0.0716	0.0729	0.0181	0.1215	0.1247	0.1072	0.2598	0.3028	0.1487	0.4707	0.5528
W-S	0.0156	0.0697	0.0698	0.0188	0.1179	0.1184	-0.0124	0.1599	0.1658	0.0026	0.2759	0.2982
Sim	-0.0155	0.0665	0.1029	-0.0162	0.1120	0.2201	-0.0341	0.1572	0.1607	-0.0218	0.2689	0.2823
$T=1000$												
Root-k	-0.0686	0.0971	0.0535	-0.1134	0.2789	0.1069	-0.3175	0.2369	0.0841	-0.3994	0.5688	0.1681
G-C	0.0356	0.0324	0.0342	0.0397	0.0541	0.0564	0.1192	0.1226	0.1301	0.1453	0.2085	0.2197
W-S	0.0365	0.0324	0.0318	0.0459	0.0554	0.0527	0.0133	0.0718	0.0756	0.0287	0.1268	0.1342
Sim	0.0036	0.0282	0.0442	0.0070	0.0483	0.0834	-0.0007	0.0680	0.0766	0.0115	0.1208	0.1305

continued

Method	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
$T=500$												
Root-k	-0.4669	0.5678	0.2195	-0.5792	1.1866	0.4389	-0.8450	1.1638	0.2880	-0.9896	1.9949	0.5761
G-C	0.0504	0.2971	0.3325	0.0746	0.5136	0.5866	0.1843	0.6975	0.7549	0.2678	1.2226	1.3603
W-S	-0.0217	0.2353	0.2453	0.0007	0.4081	0.4494	-0.0880	0.4487	0.4903	-0.0515	0.7757	0.9147
Sim	-0.0442	0.2328	0.2990	-0.0274	0.3978	0.6056	-0.0664	0.4847	0.5137	-0.0239	0.8356	0.9082
$T=1000$												
Root-k	-0.4451	0.3673	0.1069	-0.5486	0.8183	0.2139	-0.8199	0.8825	0.1403	-0.9644	1.5962	0.2807
G-C	0.0787	0.1343	0.1498	0.0955	0.2270	0.2507	0.2250	0.3690	0.3755	0.2835	0.6486	0.6632
W-S	0.0079	0.1058	0.1122	0.0282	0.1872	0.2034	-0.0583	0.2034	0.2218	-0.0357	0.3529	0.4087
Sim	-0.0025	0.1011	0.1398	0.0133	0.1798	0.2649	-0.0087	0.2170	0.2467	0.0235	0.3855	0.4247

Table A4: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.4h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.1550	0.1923	0.1061	-0.2367	0.5305	0.2122	-0.5240	0.5183	0.1669	-0.6903	1.1974	0.3338
G-C	0.0477	0.0711	0.0806	0.0560	0.1250	0.1394	0.2882	0.4914	0.5593	0.3983	0.9888	1.1183
W-S	0.0455	0.0659	0.0709	0.0596	0.1166	0.1222	0.0134	0.1689	0.1904	0.0463	0.3209	0.3637
Sim	-0.0137	0.0603	0.1004	-0.0099	0.1056	0.2191	-0.0321	0.1630	0.1866	-0.0133	0.2980	0.3359
						$T=1000$						
Root-k	-0.1521	0.1403	0.0570	-0.2310	0.4246	0.1140	-0.5204	0.4339	0.0896	-0.6835	1.0253	0.1792
G-C	0.0579	0.0385	0.0389	0.0681	0.0637	0.0655	0.2778	0.2611	0.2286	0.3516	0.4522	0.4170
W-S	0.0565	0.0367	0.0346	0.0734	0.0614	0.0581	0.0222	0.0863	0.0902	0.0470	0.1548	0.1704
Sim	-0.0030	0.0307	0.0491	0.0028	0.0514	0.1001	-0.0098	0.0804	0.0906	0.0054	0.1422	0.1596

continued

Method	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.7145	0.8124	0.2123	-0.9268	1.7687	0.4246	-1.2594	1.9530	0.2786	-1.5780	3.6402	0.5573
G-C	0.1230	0.3903	0.4758	0.1865	0.7622	0.9170	0.2453	0.7538	0.8012	0.4029	1.5126	1.7512
W-S	0.0069	0.2560	0.2884	0.0533	0.4946	0.5656	-0.0916	0.5463	0.6246	-0.0280	1.0632	1.2839
Sim	-0.0421	0.2466	0.3424	-0.0202	0.4505	0.7434	-0.0635	0.5694	0.6528	-0.0129	1.0621	1.2116
						$T=1000$						
Root-k	-0.7105	0.7044	0.1140	-0.9194	1.5472	0.2279	-1.2550	1.8076	0.1496	-1.5695	3.3420	0.2686
G-C	0.1327	0.1923	0.2091	0.1751	0.3377	0.3779	0.3256	0.4776	0.4153	0.4853	0.9736	0.8606
W-S	0.0160	0.1301	0.1361	0.0491	0.2352	0.2652	-0.0913	0.2765	0.2866	-0.0605	0.4948	0.5834
Sim	-0.0145	0.1209	0.1696	0.0048	0.2130	0.3516	-0.0212	0.2766	0.3143	0.0151	0.4975	0.5728

Table A5: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.1y_{t-1}^2 + 0.8h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	0.0109	0.1028	0.0831	0.0070	0.2943	0.1661	-0.1884	0.1876	0.1306	-0.2928	0.5099	0.2613
G-C	-0.0044	0.0813	0.0818	-0.0030	0.1737	0.1740	0.0180	0.1662	0.1672	0.0242	0.3682	0.3809
W-S	-0.0075	0.0832	0.0813	-0.0079	0.1759	0.1729	-0.0184	0.1574	0.1538	-0.0314	0.3417	0.3417
Sim	-0.0067	0.0811	0.0950	-0.0090	0.1739	0.2285	-0.0157	0.1543	0.1617	-0.0328	0.3393	0.3540
						$T=1000$						
Root-k	0.0182	0.0495	0.0388	0.0172	0.1780	0.0775	-0.1792	0.1014	0.0610	-0.2800	0.3230	0.1219
G-C	0.0055	0.0367	0.0390	0.0104	0.0782	0.0838	0.0309	0.0745	0.0792	0.0473	0.1666	0.1807
W-S	0.0009	0.0365	0.0388	0.0036	0.0776	0.0831	-0.0053	0.0690	0.0734	-0.0088	0.1515	0.1640
Sim	0.0028	0.0368	0.0464	0.0044	0.0795	0.1155	-0.0008	0.0695	0.0751	-0.0073	0.1550	0.1669

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2895	0.2741	0.1662	-0.4416	0.7194	0.3323	-0.6028	0.6029	0.2181	-0.9204	1.4810	0.4362
G-C	0.0540	0.2519	0.2527	0.0798	0.5654	0.5988	0.2730	0.6877	0.6274	0.3881	1.5960	1.5810
W-S	-0.0269	0.2143	0.2090	-0.0442	0.4660	0.4694	-0.0319	0.3645	0.3588	-0.0725	0.8212	0.8378
Sim	-0.0202	0.2117	0.2439	-0.0444	0.4656	0.5897	-0.0297	0.3627	0.3798	-0.0770	0.8245	0.8724
						$T=1000$						
Root-k	-0.2791	0.1635	0.0775	-0.4271	0.4798	0.1551	-0.5909	0.4545	0.1018	-0.9036	1.1606	0.2035
G-C	0.0663	0.1127	0.1172	0.1049	0.2564	0.2760	0.2824	0.3416	0.2901	0.4249	0.8170	0.7314
W-S	-0.0108	0.0942	0.0998	-0.0153	0.2065	0.2253	-0.0111	0.1604	0.1706	-0.0292	0.3672	0.3996
Sim	-0.0018	0.0955	0.1162	-0.0118	0.2127	0.2907	-0.0064	0.1631	0.1767	-0.0298	0.3805	0.4101

Table A6: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.8h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2107	0.1833	0.1250	-0.4060	0.6224	0.2500	-0.7258	0.7103	0.1966	-1.3422	2.3136	0.3932
G-C	0.0863	0.1554	0.1625	0.1628	0.3866	0.3752	0.6075	1.2142	1.0112	1.5996	6.4715	3.7075
W-S	0.0678	0.1431	0.1495	0.0945	0.3237	0.3209	0.0306	0.3088	0.3456	0.0573	0.8202	0.8742
Sim	-0.0010	0.1328	0.1537	-0.0045	0.3022	0.3967	-0.0041	0.3045	0.3436	-0.0158	0.7958	0.8549
						$T=1000$						
Root-k	-0.2093	0.1110	0.0585	-0.4046	0.4602	0.1170	-0.7237	0.6019	0.0920	-1.3397	2.0751	0.1840
G-C	0.0857	0.0716	0.0753	0.1686	0.1870	0.1741	0.5643	0.6909	0.4118	1.5103	4.1453	1.8562
W-S	0.0668	0.0645	0.0694	0.1002	0.1464	0.1485	0.0292	0.1397	0.1606	0.0643	0.3725	0.4050
Sim	-0.0002	0.0572	0.0742	0.0008	0.1324	0.1707	0.0003	0.1351	0.1601	-0.0009	0.3621	0.3976

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
$T=500$												
Root-k	-0.9982	1.2252	0.2501	-1.8291	3.9333	0.5001	-1.7362	3.3279	0.3282	-3.1880	10.7864	0.6564
G-C	0.2453	0.7030	0.7574	0.6614	2.9366	2.5159	0.2630	0.9006	0.8973	-0.5548	3.8417	2.6051
W-S	0.0170	0.4457	0.4985	0.0479	1.2109	1.2534	-0.0789	0.8390	0.9629	-0.0686	2.6008	2.6264
Sim	-0.0055	0.4471	0.6427	-0.0221	1.1853	2.2368	-0.0070	0.8701	1.0008	-0.0311	2.5906	2.5881
$T=1000$												
Root-k	-0.9955	1.0886	0.1170	-1.8261	3.6370	0.2341	-1.7330	3.1595	0.1536	-3.1839	10.4207	0.3072
G-C	0.2342	0.3353	0.3322	0.6133	1.4107	1.1900	0.3805	0.4990	0.3990	-0.0732	0.8876	1.1909
W-S	0.0162	0.2033	0.2319	0.0578	0.5530	0.6053	-0.0822	0.3934	0.4457	-0.0642	1.1934	1.3070
Sim	0.0010	0.1980	0.2849	-0.0017	0.5413	1.1025	0.0045	0.3937	0.4653	-0.0016	1.1907	1.2980

Table A7: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.1y_{t-1}^2 + 0.89h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
$T=500$												
Root-k	-0.0115	0.0560	0.0689	-0.0393	0.1260	0.1378	-0.2017	0.1257	0.1083	-0.3758	0.3127	0.2167
G-C	-0.0211	0.0669	0.0822	-0.0346	0.1750	0.2081	0.0039	0.1290	0.1633	0.0157	0.3603	0.4520
W-S	-0.0257	0.0668	0.0818	-0.0428	0.1747	0.2066	-0.0351	0.1221	0.1522	-0.0667	0.3286	0.4029
Sim	-0.0234	0.0675	0.0873	-0.0395	0.1768	0.2492	-0.0281	0.1234	0.1551	-0.0555	0.3339	0.4119
$T=1000$												
Root-k	0.0198	0.0330	0.0346	0.0102	0.0842	0.0692	-0.1703	0.0705	0.0544	-0.3268	0.1933	0.1089
G-C	0.0120	0.0335	0.0424	0.0203	0.0860	0.1103	0.0441	0.0693	0.0835	0.0881	0.1977	0.2352
W-S	0.0073	0.0331	0.0422	0.0114	0.0846	0.1095	0.0058	0.0611	0.0783	0.0055	0.1621	0.2121
Sim	0.0100	0.0336	0.0445	0.0155	0.0865	0.1272	0.0121	0.0612	0.0799	0.0173	0.1693	0.2177

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k = 5$			$k = 10$			$k = 5$			$k = 10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T = 500$						
Root-k	-0.2996	0.2002	0.1378	-0.5474	0.5160	0.2756	-0.5926	0.5135	0.1809	-1.0700	1.4624	0.3618
G-C	0.0402	0.1922	0.2389	0.0987	0.5776	0.7176	0.2730	0.5406	0.5452	0.5524	1.8834	1.6727
W-S	-0.0443	0.1655	0.2048	-0.0863	0.4442	0.5483	-0.0412	0.2658	0.3358	-0.0943	0.7568	0.9510
Sim	-0.0305	0.1679	0.2207	-0.0652	0.4562	0.6196	-0.0327	0.2701	0.3459	-0.0790	0.7804	0.9867
						$T = 1000$						
Root-k	-0.2670	0.1245	0.0692	-0.4964	0.3522	0.1385	-0.5656	0.4119	0.0909	-1.0291	1.2260	0.1818
G-C	0.0842	0.1087	0.1205	0.1793	0.3421	0.3620	0.3224	0.3790	0.2685	0.6703	1.4774	0.8792
W-S	0.0020	0.0827	0.1052	-0.0025	0.2216	0.2883	0.0110	0.1378	0.1713	0.0070	0.3908	0.4680
Sim	0.0143	0.0837	0.1158	0.0204	0.2362	0.3253	0.0168	0.1379	0.1770	0.0203	0.4172	0.4872

Table A8: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.05y_{t-1}^2 + 0.89h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.1132	0.0765	0.0902	-0.2594	0.1970	0.1803	-0.4373	0.3048	0.1418	-0.8930	1.0280	0.2835
G-C	0.0187	0.0835	0.1229	0.0478	0.2206	0.3259	0.1861	0.3020	0.3889	0.5968	1.5351	1.6512
W-S	0.0103	0.0815	0.1176	0.0130	0.2059	0.2989	-0.0237	0.1645	0.2394	-0.0461	0.4612	0.6861
Sim	-0.0218	0.0778	0.1148	-0.0434	0.1978	0.3194	-0.0301	0.1619	0.2431	-0.0657	0.4546	0.6858
						$T=1000$						
Root-k	-0.1040	0.0490	0.0519	-0.2450	0.1373	0.1038	-0.4406	0.2868	0.0816	-0.9030	1.0280	0.1632
G-C	0.0363	0.0491	0.0697	0.0886	0.1341	0.1864	0.2058	0.1900	0.1951	0.6565	1.2642	0.8725
W-S	0.0280	0.0479	0.0675	0.0528	0.1213	0.1740	-0.0014	0.0888	0.1327	0.0083	0.2519	0.3827
Sim	-0.0056	0.0459	0.0697	-0.0087	0.1120	0.1847	-0.0076	0.0891	0.1350	-0.0113	0.2498	0.3840

continued

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.6096	0.5333	0.1803	-1.2260	1.8457	0.3607	-1.0768	1.4609	0.2367	-2.1433	5.3391	0.4734
G-C	0.0921	0.2994	0.4251	0.2775	1.1214	1.5135	0.2730	0.6161	0.7021	0.2138	1.2817	1.6140
W-S	-0.0421	0.2317	0.3325	-0.0770	0.6711	0.9824	-0.1008	0.3977	0.5697	-0.1792	1.2675	1.8301
Sim	-0.0360	0.2321	0.3544	-0.0770	0.6750	1.0788	-0.0416	0.4118	0.6136	-0.1021	1.3057	1.8294
						$T=1000$						
Root-k	-0.6182	0.5312	0.1038	-1.2471	1.9229	0.2075	-1.1066	1.5959	0.1362	-2.1710	5.4702	0.2724
G-C	0.1220	0.1709	0.2238	0.3507	0.7290	0.7718	0.3375	0.4355	0.3927	0.4337	0.9015	0.8487
W-S	-0.0157	0.1233	0.1826	-0.0123	0.3606	0.5431	-0.0722	0.2079	0.3043	-0.1011	0.6726	0.9793
Sim	-0.0094	0.1237	0.1953	-0.0102	0.3668	0.5828	-0.0083	0.2178	0.3290	-0.0135	0.7084	1.0300

Table A9: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim$ nid(0,1) and $h_t = \omega + 0.89h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2120	0.1029	0.0798	-0.4592	0.3233	0.1596	-0.6426	0.5654	0.1255	-1.3467	2.1975	0.2509
G-C	0.0362	0.0921	0.1233	0.0945	0.2534	0.3400	0.4337	0.6265	0.5515	1.5272	5.2704	2.9383
W-S	0.0180	0.0849	0.1145	0.0126	0.2053	0.2928	-0.0266	0.1839	0.2542	-0.0508	0.5295	0.6991
Sim	-0.0326	0.0797	0.1163	-0.0608	0.1999	0.3250	-0.0458	0.1778	0.2536	-0.0907	0.5219	0.6952
						$T=1000$						
Root-k	-0.2061	0.0955	0.0566	-0.4515	0.3160	0.1132	-0.6570	0.6901	0.0890	-1.3453	2.1729	0.1780
G-C	0.0551	0.0515	0.0813	0.1396	0.1528	0.2229	0.4822	0.6784	0.3435	1.6525	5.0266	1.6592
W-S	0.0358	0.0490	0.0772	0.0510	0.1197	0.1964	-0.0031	0.0930	0.1626	0.0034	0.2669	0.3894
Sim	-0.0179	0.0481	0.0739	-0.0289	0.1152	0.2004	-0.0241	0.0940	0.1626	-0.0393	0.2653	0.3910

continued

Method	$VaR^{0.99}_{T,k}$						$ES^{0.99}_{T,k}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.8730	1.0125	0.1596	-1.8145	3.9774	0.3192	-1.4776	2.8100	0.2095	-3.0310	10.8171	0.4189
G-C	0.1408	0.4116	0.4563	0.5559	2.3048	1.7812	0.1683	0.5053	0.5514	-1.0283	3.2347	1.9745
W-S	-0.0479	0.2668	0.3608	-0.0844	0.7946	0.9891	-0.1337	0.4830	0.6028	-0.2150	1.3649	2.0271
Sim	-0.0533	0.2646	0.4266	-0.1061	0.8003	2.1882	-0.0662	0.4831	0.6252	-0.1596	1.3606	2.0313
						$T=1000$						
Root-k	-0.8986	1.2706	0.1132	-1.8223	4.0235	0.2264	-1.4904	2.9020	0.1486	-3.0539	10.9486	0.2971
G-C	0.1755	0.2347	0.3026	0.5999	1.3016	0.9695	0.2655	0.3389	0.2949	-0.7958	1.7170	0.9450
W-S	-0.0217	0.1316	0.2278	-0.0199	0.3914	0.5297	-0.1021	0.2414	0.4031	-0.1153	0.8276	1.0569
Sim	-0.0293	0.1327	0.2678	-0.0427	0.3975	1.4240	-0.0309	0.2392	0.4202	-0.0539	0.8587	1.0810

Table A10: Simulation results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where ε_t is skew-t distributed with $\lambda = -0.2$ and $\nu = 8$ and $h_t = \omega + 0.05y_{t-1}^2 + 0.8h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. The sample size and the horizon is set to 1000 observations and 5 periods, respectively. MSE is the mean square error and EAV is the average estimated asymptotic variance. The averages are for the true values.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
Average	4.8503						6.8472					
	QML			ML			QML			ML		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
Root-k	-0.1992	0.1715	0.1703	0.0356	0.1675	0.0473	-1.0146	1.1805	0.2428	-0.0787	0.3552	0.0982
G-C	-0.0602	0.1127	0.1400	0.1964	0.1640	0.0667	-0.3302	0.4342	0.4365	1.9145	6.4573	2.8491
W-S	-0.0676	0.1113	0.1275	0.0969	0.1084	0.0965	-0.5536	0.5257	0.2214	0.1070	0.3099	0.1507
Sim	-0.1077	0.1142	0.1479	0.0175	0.0922	0.1158	-0.5647	0.5364	0.2698	0.0171	0.2667	0.2046

References

Andreev, A., Kanto, A., 2005. Conditional value-at-risk estimation using non-integer values of degrees of freedom in Student's t-distribution. Journal of Risk 7, 55–61.

Baillie, R. T., Bollerslev, T., 1992. Prediction in dynamic models with time-dependent conditional variances. Journal of Econometrics 52, 91–113.

Barton, D. E., Dennis, K. E., 1952. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 39, 425–427.

Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? Journal of Finance 57, 1093 – 1111.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Wooldridge, J. M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Boudoukh, J., Richardson, M., Whitelaw, R. F., 1997. Investigation of a class of volatility estimators. Journal of Derivatives 4, 63–71.

Brännäs, K., Nordman, N., 2003. Conditional skewness modelling for stock returns. Applied Economics Letters 10, 725–728.

Breuer, T., Jandačka, M., 2007. Temporal aggregation of GARCH models: Conditional kurtosis and optimal frequency. Working paper, PPE Research Centre, University of Applied Sciences Vorarlberg.

Brummelhuis, R., Guégan, D., 2005. Multiperiod conditional distribution functions for conditionally normal GARCH(1,1) models. Journal of Applied Probability 42, 426–445.

Brummelhuis, R., Kaufmann, R., 2007. Time scaling of VAR in GARCH(1,1) and AR(1)-GARCH(1,1) processes. Journal of Risk 9, 39–94.

Chan, N. H., Deng, S.-J., Peng, L., Xia, Z., 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics 137, 556–576.

Chen, S. X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6, 87–107.

Chen, S. X., Tang, C. Y., 2005. Nonparametric inference of value-at-risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Christoffersen, P., Diebold, F. X., 2000. How relevant is volatility forecasting for financial risk management? The The Review of Economics and Statistics 82, 12–22.

Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. Journal of Risk 7, 1–28.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics 2, 84–108.

Cornish, E. A., Fisher, R. A., 1938. Moments and cumulants in the specification of distributions. Review of the International Statistical Institute 5, 307–320.

Cotter, J., 2007. Varying the VaR for unconditional and conditional environments. Journal of International Money and Finance 26, 1338–1354.

Diebold, F. X., Hickman, A., Inoue, A., Schuermann, T., 1997. Converting 1-day volatility to h-day volatility: Scaling by Root-h is worse than you think. Working paper 97-34, Wharton Financial Institutions Center.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Dowd, K., 2007. Validating multiple-period density-forecasting models. Journal of Forecasting 26, 251–270.

Duan, J.-C., Gauthier, G., Simonato, J.-G., Sasseville, C., 2006. Approximating the GJR-GARCH and EGARCH option pricing models analytically. Journal of Computational Finance 9, 41–69.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F., 1990. Stock volatility and the crash of '87: Discussion. Review of Financial Studies 3, 103–106.

Engle, R. F., Ng, V. K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Giamouridis, D., 2006. Estimation risk in financial risk management: A correction. Journal of Risk 8, 121–125.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gradshteyn, I. S., Ryzhik, I. M., 1994. Tables of Integrals, Series and Products, 5th Edition. Academic Press, New York.

Hafner, C. M., 2003. Fourth moment structure of multivariate GARCH models. Journal of Financial Econometrics 1, 26–54.

Hafner, C. M., 2008. Temporal aggregation of multivariate GARCH processes. Journal of Econometrics 142, 467–483.

Hansen, B. E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. Journal of Econometrics 135, 377–398.

Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34, 465–487.

He, C., Teräsvirta, T., 1999. Properties of moments of a family of GARCH processes. Journal of Econometrics 92, 173–192.

Jondeau, E., Rockinger, M., 2001. Gram-Charlier densities. Journal of Economic Dynamics and Control 25, 1457–1483.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Jondeau, E., Rockinger, M., 2006. The Copula-GARCH model of conditional dependencies: An international stock market application. Journal of International Money and Finance 25, 827–853.

Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill, New York.

Kaibila, P., He, Z., 2004. The adjustment of prediction intervals to account for errors in parameter estimation. Journal of Time Series Analysis 25, 351–358.

Lan, Y., Hu, Z. M., Johnson, J., 2007. Can value-at-risk be made more valuable? Resampling techniques for risk uncertainty. Working paper, Financial Studies Discipline, UWA Business School, The University of Western Australia.

Ling, S., McAleer, M., 2002. Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106, 109–117.

Manistre, B. J., Hancock, G. H., 2005. Variance of the CTE estimator. North American Actuarial Journal 9, 129–156.

Marcellino, M., Stock, J. H., Watson, M. W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135, 499–526.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271–300.

Pesaran, M. H., Pick, A., Timmermann, A., 2009. Variable selection and inference for multi-period forecasting problems. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 2007. Numerical Recipes: The Art of Scientific Computing, 3rd Edition. Cambridge University Press, New York.

Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115–129.

Schmidt, P., 1974. The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. Econometrica 42, 303–309.

Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies 62, 639–661.

Taylor, J. W., 1999. A quantile regression approach to estimating the distribution of multiperiod returns. Journal of Derivatives 7, 64–78.

Taylor, J. W., 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 19, 299–311.

Theodossiou, P., 1998. Financial data and the skewed generalized t distribution. Management Science 44, 1650–1661.

Wong, C.-M., So, M. K. P., 2003. On conditional moments of GARCH models, with applications to multiple period value at risk estimation. Statistica Sinica 13, 1015–1044.

Wong, C.-M., So, M. K. P., 2007. On the estimation of multiple period expected shortfall and median shortfall using GARCH models. Working paper, Hong Kong University of Science and Technology.

Value at Risk for Large Portfolios

Carl Lönnbark*, Ulf Holmberg and Kurt Brännäs*
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
E-mail: carl.lonnbark@econ.umu.se

Abstract

We argue that the practise used in the valuation of the portfolio is important for the calculation of the Value at Risk. In particular, when liquidating a large portfolio the seller may not face horizontal demand curves. We propose a partially new approach for incorporating this fact in the Value at Risk and in an empirical illustration we compare it to a competing approach. We find substantial differences.

Key Words: Demand, Supply, Liquidity Risk, Limit Order Book, Bank, Sweden.
JEL Classification: C22, C51, C53, D40, G00, G10.

*The financial support from the Jan Wallander and Tom Hedelius Foundation and the Tore Browaldh Foundation is gratefully acknowledged.

1 Introduction

In this paper we address the question of how to properly assess the risk in large financial portfolios. In risk assessment it is usually assumed that the entire position can be sold at the market price (or mid-price), though one realizes that this can be a quite misleading valuation approach. The reason is that for large enough positions the seller (buyer) of an asset does not face a horizontal demand (supply) curve. Thus, there is an element of liquidity risk involved (see Malz, 2003, for a general discussion of liquidity risk) and this should reasonably be taken into account in risk assessment.

Here, the primary focus is on incorporating the liquidity risk in the Value at Risk (VaR) measure, which is the standard way of quantifying the risk of adverse price movements. VaR is defined as the maximum potential portfolio loss that will not be exceeded over a given time horizon for some small probability (see Jorion, 2007, for a survey). We emphasize, as argued by François-Heude and Van Wynendaele (2001) and others, that it is implicitly assumed that the liquidation occurs in one block at the end of the predefined holding period when calculating the VaR. The question of how to incorporate the liquidity risk into the VaR is a relatively old one and several alternative approaches have been proposed. Bangia, Diebold, Schuermann, and Stroughair (1999) were the first to account for it, with their spread based alternative. Ernst, Stange, and Kaserer (2009) evaluates some measures empirically.

Our proposed approach relies on essentially the same idea as in Giot and Grammig (2006) (GG hereafter). Rather than the mid-price at the end of the horizon they consider the average price per share that would be obtained upon immediate liquidation. Their VaR is volume dependent and it is based on the difference between the mid-price at the beginning of the horizon and the average price at the end of it. We argue that the relevant initial price is not the mid-price, but that the portfolio should be valued at the average price in the beginning of the period as well. We have assets traded on an order driven markets with a visible limit order book (LOB) (e.g., Gourieroux and Jasiak, 2001, ch. 14) in mind and the context is intra-day. Though frequently used on a (at least) daily basis, intra-day VaR's are of interest as well. For example, Dionne, Duchesne, and Pacurar (2008) argue that the investment horizon for very active agents on the market is typically less then one day.

When it comes to the modelling of the dynamics of the average prices the literature is quite scarce. The model employed in GG is of AR-GARCH type and it is essentially univariate. Other previous attempts include Gourieroux, Le Fol, and Meyer (1998) and Bowsher (2004). The former consider a factor model in transaction time, while the latter proposes a functional signal plus noise time series model in calender time. Our framework shares features with all three approaches and the resulting multivariate model allows for spatial (in the volume dimension) as well as serial correlation in the time dimension.

The paper is organized as follows. In Section 2 our VaR framework is presented. Section 3 gives some descriptive statistics of our data set consisting of high-frequency observations on the limit order book of Swedish banking stocks. In Section 4 we propose a time series model for the dynamics of the limit order book. Section 5 contains some empirical results including a comparison with the competing approach of GG.

2 The liquidity adjusted VaR

The object of interest is the conditional VaR for the horizon T to $T+h$ for a univariate portfolio consisting of v_T shares of a financial asset. We will consider VaR's for both long and short portfolios. For the latter we borrow shares today and agree to return them at some future date. Thus, in that case v_T is negative. We do not allow for portfolio updating, so that $v_T = v_{T+i}$, $i = 1, ..., h$, and we denote the value of the portfolio at time point $t = T, ..., T+h$ by V_t. Following Gourieroux and Jasiak (2001, ch. 16) the VaR for the position v_T satisfies

$$\Pr\{V_{T+h} - V_T < -VaR_{T,h}^{1-\alpha} \mid \mathcal{F}_T\} = \alpha, \tag{1}$$

where $\mathcal{F}_T$ is the information available at time T. That is, with the (small) probability α the change in the value of the portfolio is less than $-VaR^{1-\alpha}_{T,h}$. In anticipation of what follows we note that the VaR depends on how we compute the values V_{T+h} and V_T. The approach typically adopted in the literature is to assume that the entire portfolio can be sold at one and the same price, e.g., the mid-price, $\tilde{P}_t$, $t = T, ..., T+h$ (say). This implies that the portfolio values V_T and V_{T+h} in (1) are approximated by $\tilde{V}_T = \tilde{P}_T v_T$ and $\tilde{V}_{T+h} = \tilde{P}_{T+h} v_T$, respectively. The corresponding approximative VaR for a long position then satisfies

$$\begin{aligned}\Pr\{\tilde{V}_{T+h} - \tilde{V}_T < -\widetilde{VaR}^{1-\alpha}_{T,h} \mid \mathcal{F}_T\} = \\ \Pr\{(\tilde{P}_{T+h} - \tilde{P}_T)v_T < -\widetilde{VaR}^{1-\alpha}_{T,h} \mid \mathcal{F}_T\}. \qquad (2)\end{aligned}$$

For a short position the expression becomes $\Pr[(\tilde{P}_{T+h} - \tilde{P}_T)v_T > \widetilde{VaR}^{1-\alpha}_{T,h} \mid \mathcal{F}_T]$. The discussion below is for a long position, but it applies analogously for a short one. Now, for relatively small positions we expect the VaR as defined by (2) to provide a reasonable approximation. However, as argued in the introduction $\tilde{V}_T$ does not in general give the correct value of the portfolio. For example, assume that our position consists of 1000 shares and that at time $T+h$, 500 shares are demanded at the price 2 at the first level of the bid-side of the LOB, and that 1000 shares are demanded at price 1 at the second level. Whereas a marking to the mid-price approach would assign a value of, at least, 2000 we would actually obtain $500 \times 2 + (1000 - 500) \times 1 = 1500$ upon immediate liquidation. The average price per unit of sold volume for this transaction is 1.5 and it appears that this is the fair price to replace for $\tilde{P}_{T+h}$ in (2).

Generalizing, we define $\bar{P}_t(v)$ as the average price as a function of the volume, i.e. the average price per unit of volume that would result from immediately executing a market order of v shares. In the sequel we let superscripts a and b indicate whether the average price is for the ask or the bid side of the LOB. Figure 1 shows demand and supply schedules along with the corresponding average price curves for an observation of one of the stocks (SWB) in our data set.

The question is then how to properly use $\bar{P}_t(v)$ to compute the relevant change in value and this is where we differ from GG. They consider a one-period setting and in their view the relevant change

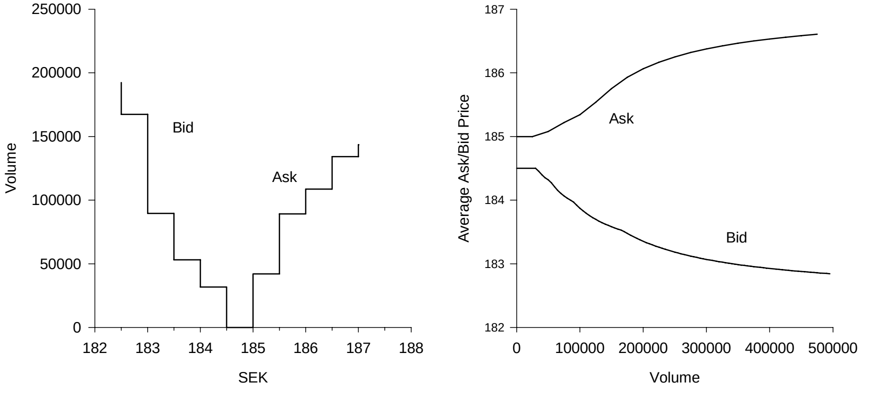

Figure 1: Supply and demand schedules (left) and average price curves (right) in SWB, August 1 at 10AM.

in the value of a position of size v_T is given by $\bar{P}^b_{T+1}(v_T)v_T - \tilde{P}_T v_T$, where $\tilde{P}_T = \left[\bar{P}^a_T(1) + \bar{P}^b_T(1)\right]/2$. They specify the dynamics of the log-returns, $p^{GG,v}_t = \ln(\bar{P}_t(v_T)/\tilde{P}_{t-1})$, on the location-scale form $p^{GG,v}_t = \mu^{GG,v}_t + \sigma^{GG,v}_t \varepsilon^{GG,v}_t$, where μ^{GG}_t and σ^{GG}_t are the conditional mean and standard deviation of $p^{GG,v}_t$, respectively, and $\varepsilon^{GG,v}_t$ is an iid random variable with zero mean and unit variance. Their VaR is[1]

$$VaR^{GG,1-\alpha}_{T,1} = -\tilde{P}_T v_T(\exp(\mu^{GG,v}_{T+1} + \sigma^{GG,v}_{T+1} q^{\eta}_{\alpha}) - 1),$$

where q^{η}_{α} is the αth quantile in the Student's t distribution with η degrees of freedom.

We argue that with the same motivation as we value the portfolio at the average price at the end of the period, we should also value it at the average price in the beginning of it. Thus the relevant one-period change in value is $\bar{P}^b_{T+1}(v_T)v_T - \bar{P}^b_T(v_T)v_T$. With the corresponding log-return dynamics

$$p^v_t = \mu^v_t + \sigma^v_t \varepsilon^v_t,$$

[1] Actually, their VaR is the quantile of the distribution of the log-returns, but this is the implication for the VaR definition we use.

our VaR alternative is

$$VaR_{T,1}^{1-\alpha} = -\bar{P}_T^b(v_T)v_T(\exp(\mu_{T+1}^v + \sigma_{T+1}^v q_\alpha) - 1), \tag{3}$$

where q_α is the αth quantile of some suitable distribution.

For a horizon of h periods the VaR satisfies $\Pr\{[\bar{P}_{T+h}(v_T) - \bar{P}_T(v_T)]v_T \leq -VaR_{T,h}^{1-\alpha} \mid \mathcal{F}_T\}$. However, the dynamics of the h-period returns do not follow easily from that of the one-period returns (cf. Lönnbark, 2009). Note also that our VaR and the VaR in Giot and Grammig (2006) are related by

$$VaR_{T,h}^{1-\alpha} = VaR_{T,h}^{GG,1-\alpha} - (\bar{P}_T(v) - \tilde{P}_T)v_T.$$

Hence, given one of the VaR's it is possible to obtain the other through an additive transformation that is known at time T. Note also that the difference between the two measures grows with an increasing volume.

The VaR in (3) implicitly assumes that we own the portfolio at T. If it is to be purchased at T we use $\bar{P}_T^a(v_T)$ for the initial price and the VaR becomes

$$VaR_{T,1}^{1-\alpha} = \bar{P}_T^a(v_T)v_T - \bar{P}_T^b(v_T)v_T\exp(\mu_{T+1}^v + \sigma_{T+1}^v q_\alpha).$$

We end this section by giving the corresponding VaR's for a short position. They are, respectively, given by

$$VaR_{T,1}^{1-\alpha} = \bar{P}_T^a(v_T)v_T(\exp(\mu_{T+1}^v + \sigma_{T+1}^v q_{1-\alpha}) - 1), \tag{4}$$

$$VaR_{T,1}^{1-\alpha} = \bar{P}_T^a(v_T)v_T\exp(\mu_{T+1}^v + \sigma_{T+1}^v q_{1-\alpha}) - \bar{P}_T^b(v_T)v_T. \tag{5}$$

Note that $q_{1-\alpha}$ instead of q_α appears in (4) and (5).

3 Data and descriptives

Our dataset consists of time series for the four largest banks in Sweden (Nordea NRD, Skandinaviska Enskilda Banken SEB, Handelsbanken SHB, Swedbank SWB[2]) and covers the period May 3 – August 8, 2005.[3]

[2] Föreningssparbanken in the sample period.

[3] For technical reasons the period June 7–10 is missing for all banks, and additionally May 27–June 1 for SWB and NRD.

Table 1: Descriptive statistics for the number of traded shares and closing prices in individual transactions for the four banks in the first trading month.

	Nr of Traded Shares			Closing Price		
Bank	Mean	StDev	Max	Mean	StDev	n
NRD	9921.6	95424.5	7 383 816	67.7	0.42	19026
SEB	5341.5	$1.12{\cdot}10^5$	12 946 377	127.7	1.81	14325
SHB	3963.1	33227.2	1 831 705	161.1	2.40	10445
SWB	3644.2	22917.7	1 299 919	171.4	2.56	11468

Table 1 gives a few descriptive statistics for the trading patterns in the four banking stocks for the first trading month (21 days) of the data. The number of traded shares distributions are quite skewed with a long upper tail and the largest transactions in each month are quite large. The largest transaction was in SEB and amounted to about 1653 million SEK using the average price. This corresponds to about 17 percent of total transactions during the month. For the other stocks the corresponding percentages are about 4 percent. Trading is most frequent in NRD with about 900 daily transactions or about 2 per minute.

The sampling frequency is chosen to be 30 minutes, such that the records immediately preceding the given half-hour are chosen. The daily records cover 1000–1700, i.e. there are 15 observations during the day and the total time series length is $T = 936$ for SEB and SHB and $T = 861$ for NRD and SWB.

For the empirical modelling we can obtain time series of average prices for any chosen volume level. For the analyses reported later we have chosen five volume levels $v = 1, 100000(50000)300000$ and all results are based on log-returns $p_t^v = \ln(\bar{P}_t^v) - \ln(\bar{P}_{t-1}^v)$. As an illustration of the spatial/volume correlations within stocks we consider log-returns for the ask side of SHB, cf. Table 2. As expected from the smoothness of the average curve in Figure 1, we find that correlations between log-returns at the different volume levels are close to 1. Obviously, the correlations are weaker for lagged log-returns. The autocorrelation function closely matches the cross correlation, except for the first volume level.

Table 2: Cross correlations for log-returns (ask) in SHB across volume levels with $v = 200000$ as a base.

	Volumes (thousands)					
Lag	$1 \cdot 10^{-3}$	100	150	200	250	300
0	0.77	0.94	0.99	1.00	0.99	0.97
1	-0.01	-0.00	-0.03	-0.04	-0.04	-0.04
2	-0.02	-0.13	-0.13	-0.13	-0.13	-0.13
3	-0.07	-0.05	-0.06	-0.07	-0.07	-0.07
4	-0.03	0.00	0.00	0.01	0.01	0.01

Table 3: Parameter estimates and descriptive statistics for MA(1) models and their residuals of the ask/bid (a and b) average log-return series of the four banks at volume level $v = 200000$. p-values are used for the Ljung-Box statistics, LB.

Bank		MA(1)	t	LB_{10}	LB^2_{10}	Skew	Kurt
NRD	a	0.088	2.59	0.86	0.02	0.26	3.60
	b	0.141	4.15	0.95	0.03	-0.04	3.58
SEB	a	-0.030	-0.90	0.83	0.00	2.10	7.97
	b	-0.010	-0.32	0.82	0.00	-0.38	15.7
SHB	a	0.044	1.34	0.05	0.00	0.76	6.73
	b	0.063	1.94	0.60	0.00	-0.08	5.56
SWB	a	0.088	2.60	0.31	0.63	1.11	8.39
	b	0.018	0.54	0.63	0.00	-0.64	14.0

Based on the SWB series the autocorrelation functions suggest that MA(1) models will account for most of the serial correlation in the time series. Table 3 gives estimated models and some descriptive statistics for the residuals of the models. In all but one case there is significant autocorrelation in squared residuals, suggesting that ARCH effects are of major importance. For the ask series there is positive skewness and weak but negative for the bid series. For most series there is substantial kurtosis.

4 A time series model for the average price curves

We specify the dynamics of the average price curves in terms of log returns. Stock prices are widely taken to be random walks with drift and for returns various autoregressive and/or moving average extensions of the basic model seem to empirically surface. Based on some initial specification searches on the SWB stock we take as a reasonable model

$$
\begin{aligned}
p_t^{v_1} &= \alpha_{v_1} + \boldsymbol{\beta}'\mathbf{d}_t + \varepsilon_t^{v_1} + \theta_0 \varepsilon_{t-1}^{v_1} \\
p_t^{v_i} &= \alpha_{v_i} + \boldsymbol{\beta}'\mathbf{d}_t + \gamma_{v_i} p_{t-1}^{v_{i-1}} + \varepsilon_t^{v_i} + \theta_{v_i}\varepsilon_{t-1}^{v_i}, \qquad i = 2, \ldots, m,
\end{aligned}
$$

where $p_t^{v_i} = \log[\bar{P}_t(v_i)] - \log[\bar{P}_{t-1}(v_i)]$. The parameters γ_{v_i} and θ_{v_i} are volume dependent; $\gamma_{v_i} = \gamma_0 + \gamma_1 v_{i-1}$ and $\theta_{v_i} = \theta_0 + \theta_1 v_i$, $i = 2, \ldots, m$. The $\mathbf{d}_t$ is a vector of dummy variables to catch overnight impacts on the first observation of the day and time of day effects. In addition, the models of different volume levels may be correlated such that $E(\varepsilon_{t,v_i}\varepsilon_{s,v_j}) \neq 0$, for all v_i, v_j and also for $t \neq s$.

For all volume levels, $\mathbf{v} = \{v_1, v_2, ..., v_m\}$, we write

$$
\begin{pmatrix} p_t^{v_1} \\ \vdots \\ p_t^{v_m} \end{pmatrix} = \begin{pmatrix} \alpha_{v_1} \\ \vdots \\ \alpha_{v_m} \end{pmatrix} + \boldsymbol{\beta}'\mathbf{d}_t\boldsymbol{\iota} + \begin{pmatrix} 0 & & \cdots & & 0 \\ \gamma_{v_2} & & & & \\ 0 & \gamma_{v3} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & 0 & \gamma_{v_m} & 0 \end{pmatrix}
$$

$$
\times \begin{pmatrix} p_{t-1}^{v_1} \\ \vdots \\ p_{t-1}^{v_m} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{v_1} \\ \vdots \\ \varepsilon_t^{v_m} \end{pmatrix} + \begin{pmatrix} \theta_{v_1} & 0 & \cdots & 0 \\ 0 & \theta_{v_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \theta_{v_m} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1}^{v_1} \\ \vdots \\ \varepsilon_{t-1}^{v_m} \end{pmatrix}
$$

or compactly

$$\mathbf{p}_t^{\mathbf{v}} = \boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{d}_t\boldsymbol{\iota} + \boldsymbol{\Gamma}_{\mathbf{v}}\mathbf{p}_{t-1}^{\mathbf{v}} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_{\mathbf{v}}\boldsymbol{\varepsilon}_{t-1}, \tag{6}$$

where $\boldsymbol{\iota}$ is a vector of ones and $\boldsymbol{\varepsilon}_t$ has zero mean and conditional covariance matrix $\boldsymbol{\Sigma}_t$. Thus, the model is of VARMAX type and has both a time series and volume/spatial dimension. The $\boldsymbol{\Sigma}_t$ may contain nonzero off-diagonal elements and is also indexed by t to allow for ARCH-effects. For the conditional variances we employ a version of the asymmetric GARCH specification of Glosten, Jagannathan, and Runkle (1993)

$$h_t^{v_i} = \omega_{v_i} + \delta h_t^{v_i} + \eta(\varepsilon_{t-1}^{v_i})^2 + \lambda(\varepsilon_{t-1}^{v_i})^2\mathbf{1}(\varepsilon_{t-1}^{v_i} < 0), \tag{7}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Note that ω_{v_i} is the only parameter that changes across v_i. As a full baseline model for $\boldsymbol{\Sigma}_t$ we consider (7) together with constant off-diagonal elements

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Omega} + \delta \operatorname{diag}(\mathbf{h}_t^{\mathbf{v}}) + \eta \operatorname{diag}(\boldsymbol{\varepsilon}_{t-1}^{2,\mathbf{v}}) + \lambda \operatorname{diag}(\boldsymbol{\varepsilon}_{t-1}^{2-,\mathbf{v}}),$$

where $\mathbf{h}_t^{\mathbf{v}}$, $\boldsymbol{\varepsilon}_t^{2,\mathbf{v}}$ and $\boldsymbol{\varepsilon}_t^{2-,\mathbf{v}}$ have elements $h_t^{v_i}$, $(\varepsilon_t^{v_i})^2$ and $(\varepsilon_t^{v_i})^2\mathbf{1}(\varepsilon_t^{v_i} < 0)$, $i = 1, ..., m$, respectively. The diag$(\cdot)$ operator returns a matrix with the vector argument on the diagonal and zeros elsewhere. Hence, the conditional expectation and the conditional variance of the log returns are, respectively, given by

$$\begin{aligned} E(\mathbf{p}_t^{\mathbf{v}}|\mathcal{F}_{t-1}) &= \boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{d}_t\boldsymbol{\iota} + \boldsymbol{\Gamma}_{\mathbf{v}}\mathbf{p}_{t-1}^{\mathbf{v}} + \boldsymbol{\Theta}_{\mathbf{v}}\boldsymbol{\varepsilon}_{t-1} \\ V(\mathbf{p}_t^{\mathbf{v}}|\mathcal{F}_{t-1}) &= \boldsymbol{\Sigma}_t. \end{aligned} \tag{8}$$

These expression are useful both for estimation and forecasting over time. From (6) it is straightforward to obtain the corresponding price levels as $\bar{P}_t^{v_i} = \bar{P}_{t-1}^{v_i}\exp(p_t^{v_i})$, $i = 1, ..., m$. The conditional expectation and variance of $\bar{P}_t^{v_i}$ may be obtained by taking first order expansions of the exponential function and (8).

With respect to the spatial aspects of the model note that this is an unusual context of observation availability for all volume levels. However, for low levels the volume curves are typically flat and for very large levels linear. Therefore, it appears reasonable to focus the modelling exercise on the intermediate levels, where the curvature is most pronounced. The way we choose v and m in the estimation phase impacts the precision of the estimates, but as our model is not able to predict in the volume direction, the choice is also practically related to the model's end use for VaR calculations.

4.1 Estimation

When it comes to predicting the VaR we use a multivariate version of a popular methodology known as filtered historical simulation (FHS) in the literature (e.g., Christoffersen, 2009). To explain the approach we first collect all model parameters in the vector $\boldsymbol{\psi}$ and consider the prediction error $\mathbf{e}_t = \mathbf{p}_t^{\mathbf{v}} - E_{\boldsymbol{\psi}}(\mathbf{p}_t^{\mathbf{v}}|\mathcal{F}_{t-1})$, where we subindex the expectation operator to emphasize that it is to be taken under $\boldsymbol{\psi}$. Assuming that the standardized prediction errors $\tilde{\mathbf{e}}_t = (\boldsymbol{\Sigma}_t^{1/2})^{-1}\mathbf{e}_t$, $t = 1, ..., T$, is an iid sequence we may approximate the conditional distribution of $\mathbf{p}_{T+1}^{\mathbf{v}}$ with the sequence $\mathbf{p}_{T+1,j}^{\mathbf{v},*} = E_{\boldsymbol{\psi}}(\mathbf{p}_{T+1}^{\mathbf{v}}|\mathcal{F}_T) + \boldsymbol{\Sigma}_{T+1}^{1/2}\tilde{\mathbf{e}}_j$, $j = 1, ..., T$. The predictors of the one-period VaR's are then trivially obtained from suitable empirical quantiles of the $\mathbf{p}_{T+1,j}^{\mathbf{v},*}$ sequences.

The FHS is a two-step procedure that in the first step estimates the underlying model parameters employing some estimator, $\hat{\boldsymbol{\psi}}$. In the second step it filters out the $\tilde{\mathbf{e}}_t$ sequence.

A natural choice for $\hat{\boldsymbol{\psi}}$ is the quasi maximum likelihood estimator. Given observations up til time T it involves finding the $\boldsymbol{\psi}$ that maximizes the log-likelihood function

$$\ln L = -\frac{1}{2}\sum\nolimits_{t=2}^{T}(\ln|\boldsymbol{\Sigma}_t| - \mathbf{e}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{e}_t).$$

For practical estimation we use the RATS 6.0 package and employ robust standard errors.

5 Empirical results

The empirical results are summarized in terms of VaR measures in Table 4 for the case when we own the portfolio at the horizon origin. Parameter estimates may be found in Table 5. The measures are calculated for the

Table 4: VaR estimates for $\alpha = 0.01$.

	NRD		SEB		SHB		SWB	
Volume	Short	Long	Short	Long	Short	Long	Short	Long
1	0.59	0.38	1.29	1.33	1.30	0.53	1.63	0.86
100 000	57 698	37 733	136 107	119 240	125 733	74 944	175 379	92 403
150 000	86 138	48 532	225 120	161 578	198 631	116 893	281 068	174 078
200 000	113 910	61 606	307 464	212 012	313 717	182 451	416 784	222 729
250 000	139 688	74 507	414 339	286 324	463 808	248 137	528 713	299 420
300 000	166 669	84 207	517 234	366 692	591 360	341 902	606 274	382 042

first post sample time period, i.e. 5PM of August 8 to 10AM of August 9, 2005. The numbers reported for a short position are throughout larger than the ones for the corresponding long position. This is a consequence of, at least, the asymmetry in average cost curves.

Figure 2 gives VaR's per share for SWB. With some exceptions, there is a modest growth in all measures. If we take the view that we own the portfolio at the horizon origin, our VaR's are smaller than those calculated as in GG. If the portfolio is to be purchased, they are larger. Noteworthy is also that for the latter view our VaR's rise more sharply with volume. There is a growing difference between our VaR's and the ones as in GG, starting from one half of a tick (0.25 SEK) at volume 1 to exceeding 2 ticks for the largest position of $v = 300000$ shares. Obviously, these differences will have substantial consequences for how to set the required capital for large financial institutions.

Table 5: Parameter estimates. Bold facing indicates significance at 5 percent level. The Ljung-Box statistics are evaluated at the volume level $v = 200000$.

	NRD		SEB		SHB		SWB	
Param	Ask	Bid	Ask	Bid	Ask	Bid	Ask	Bid
α_0	$-2.75 \cdot 10^{-5}$	$-2.20 \cdot 10^{-5}$	$3.84 \cdot 10^{-6}$	$1.51 \cdot 10^{-4}$	$-1.68 \cdot 10^{-5}$	$4.72 \cdot 10^{-5}$	$-6.72 \cdot 10^{-5}$	$5.43 \cdot 10^{-5}$
α_1	$-2.97 \cdot 10^{-5}$	$-3.21 \cdot 10^{-5}$	$1.10 \cdot 10^{-4}$	$\mathbf{1.62 \cdot 10^{-4}}$	$-3.88 \cdot 10^{-5}$	$4.43 \cdot 10^{-5}$	$-9.19 \cdot 10^{-5}$	$1.42 \cdot 10^{-5}$
α_2	$-3.27 \cdot 10^{-5}$	$-8.15 \cdot 10^{-5}$	$1.31 \cdot 10^{-4}$	$1.58 \cdot 10^{-4}$	$-3.47 \cdot 10^{-5}$	$3.89 \cdot 10^{-5}$	$-7.96 \cdot 10^{-5}$	$-1.70 \cdot 10^{-5}$
α_3	$-3.49 \cdot 10^{-5}$	$-3.93 \cdot 10^{-5}$	$1.40 \cdot 10^{-4}$	$1.48 \cdot 10^{-4}$	$-3.05 \cdot 10^{-5}$	$3.73 \cdot 10^{-5}$	$-6.93 \cdot 10^{-5}$	$\mathbf{-3.84 \cdot 10^{-5}}$
α_4	$-3.74 \cdot 10^{-5}$	$-5.16 \cdot 10^{-5}$	$1.58 \cdot 10^{-4}$	$1.55 \cdot 10^{-4}$	$-2.95 \cdot 10^{-5}$	$3.03 \cdot 10^{-5}$	$-6.21 \cdot 10^{-5}$	$\mathbf{-5.48 \cdot 10^{-5}}$
α_5	$-4.06 \cdot 10^{-5}$	$-6.06 \cdot 10^{-5}$	$1.76 \cdot 10^{-4}$	$1.72 \cdot 10^{-4}$	$-2.97 \cdot 10^{-5}$	$2.45 \cdot 10^{-5}$	$-5.79 \cdot 10^{-5}$	$\mathbf{-6.95 \cdot 10^{-5}}$
β_0	$\mathbf{2.27 \cdot 10^{-3}}$	$\mathbf{2.07 \cdot 10^{-3}}$	$\mathbf{1.96 \cdot 10^{-3}}$	$\mathbf{-5.55 \cdot 10^{-4}}$	$\mathbf{2.70 \cdot 10^{-3}}$	$5.55 \cdot 10^{-4}$	$\mathbf{4.13 \cdot 10^{-3}}$	$7.36 \cdot 10^{-5}$
β_{on}	$1.67 \cdot 10^{-6}$	$7.19 \cdot 10^{-6}$	$-3.88 \cdot 10^{-4}$	$-2.97 \cdot 10^{-5}$	$\mathbf{-3.68 \cdot 10^{-4}}$	$-1.59 \cdot 10^{-5}$	$\mathbf{-4.25 \cdot 10^{-4}}$	$1.09 \cdot 10^{-4}$
β_m	$-9.76 \cdot 10^{-5}$	$-6.64 \cdot 10^{-5}$	$-7.18 \cdot 10^{-5}$	$\mathbf{-8.49 \cdot 10^{-5}}$	$7.44 \cdot 10^{-5}$	$3.51 \cdot 10^{-6}$	$1.39 \cdot 10^{-4}$	$\mathbf{1.59 \cdot 10^{-4}}$
γ_0	0.0407	$-5.35 \cdot 10^{-3}$	-0.0136	**0.1071**	0.0217	**0.1867**	0.0331	**0.2763**
γ_1	$2.22 \cdot 10^{-7}$	$\mathbf{-2.15 \cdot 10^{-7}}$	$-7.88 \cdot 10^{-7}$	$\mathbf{-3.65 \cdot 10^{-7}}$	$\mathbf{6.92 \cdot 10^{-7}}$	$\mathbf{1.01 \cdot 10^{-6}}$	$-4.08 \cdot 10^{-7}$	$\mathbf{2.55 \cdot 10^{-6}}$
θ_0	**-0.1178**	**-0.0975**	-0.0612	**-0.1033**	0.0520	**-0.1818**	**-0.1440**	**-0.0818**
θ_1	$\mathbf{-1.13 \cdot 10^{-7}}$	$-6.49 \cdot 10^{-8}$	$8.32 \cdot 10^{-7}$	$\mathbf{3.81 \cdot 10^{-7}}$	$\mathbf{-1.41 \cdot 10^{-6}}$	$\mathbf{-1.26 \cdot 10^{-6}}$	$\mathbf{3.21 \cdot 10^{-7}}$	$\mathbf{-2.80 \cdot 10^{-6}}$
LB_{10}	6.01	4.12	4.39	9.63	20.3	3.90	11.6	18.8

continued

Param	NRD Ask	NRD Bid	SEB Ask	SEB Bid	SHB Ask	SHB Bid	SWB Ask	SWB Bid
ω_0	$5.35 \cdot 10^{-6}$	$\mathbf{5.51 \cdot 10^{-6}}$	$\mathbf{3.84 \cdot 10^{-6}}$	$\mathbf{6.03 \cdot 10^{-6}}$	$\mathbf{3.47 \cdot 10^{-6}}$	$\mathbf{3.39 \cdot 10^{-6}}$	$\mathbf{3.74 \cdot 10^{-6}}$	$\mathbf{3.72 \cdot 10^{-6}}$
ω_1	$4.91 \cdot 10^{-6}$	$\mathbf{5.56 \cdot 10^{-6}}$	$\mathbf{3.71 \cdot 10^{-6}}$	$\mathbf{5.69 \cdot 10^{-6}}$	$\mathbf{2.98 \cdot 10^{-6}}$	$\mathbf{2.78 \cdot 10^{-6}}$	$\mathbf{3.67 \cdot 10^{-6}}$	$\mathbf{3.38 \cdot 10^{-6}}$
ω_2	$4.80 \cdot 10^{-6}$	$\mathbf{5.77 \cdot 10^{-6}}$	$\mathbf{3.83 \cdot 10^{-6}}$	$\mathbf{6.11 \cdot 10^{-6}}$	$\mathbf{3.23 \cdot 10^{-6}}$	$\mathbf{2.92 \cdot 10^{-6}}$	$\mathbf{4.25 \cdot 10^{-6}}$	$\mathbf{3.79 \cdot 10^{-6}}$
ω_3	$4.72 \cdot 10^{-6}$	$\mathbf{5.75 \cdot 10^{-6}}$	$\mathbf{4.19 \cdot 10^{-6}}$	$\mathbf{6.90 \cdot 10^{-6}}$	$\mathbf{3.57 \cdot 10^{-6}}$	$\mathbf{3.10 \cdot 10^{-6}}$	$\mathbf{4.83 \cdot 10^{-6}}$	$\mathbf{4.45 \cdot 10^{-6}}$
ω_4	$4.63 \cdot 10^{-6}$	$\mathbf{5.54 \cdot 10^{-6}}$	$\mathbf{4.50 \cdot 10^{-6}}$	$\mathbf{7.51 \cdot 10^{-6}}$	$\mathbf{3.84 \cdot 10^{-6}}$	$\mathbf{3.39 \cdot 10^{-6}}$	$\mathbf{5.03 \cdot 10^{-6}}$	$\mathbf{5.01 \cdot 10^{-6}}$
ω_5	$4.55 \cdot 10^{-6}$	$\mathbf{5.64 \cdot 10^{-6}}$	$\mathbf{4.81 \cdot 10^{-6}}$	$\mathbf{7.84 \cdot 10^{-6}}$	$\mathbf{4.06 \cdot 10^{-6}}$	$\mathbf{3.71 \cdot 10^{-6}}$	$\mathbf{5.03 \cdot 10^{-6}}$	$\mathbf{5.34 \cdot 10^{-6}}$
δ	$1.75 \cdot 10^{-4}$	$-3.16 \cdot 10^{-4}$	**0.3108**	**0.0159**	$\mathbf{1.94 \cdot 10^{-3}}$	$-1.02 \cdot 10^{-4}$	$\mathbf{1.51 \cdot 10^{-3}}$	$-6.27 \cdot 10^{-5}$
η	0.0348	**0.1069**	0.1507	**0.3102**	**0.1421**	0.0786	**0.1488**	0.1576
λ	0.0335	**0.0730**	0.0935	**0.1108**	$-5.07 \cdot 10^{-3}$	0.1801	-0.0996	0.0877
LB^2_{10}	2.53	3.19	7.56	36.3	10.1	44.8	6.0	3.50
T	861	861	936	936	936	936	861	861

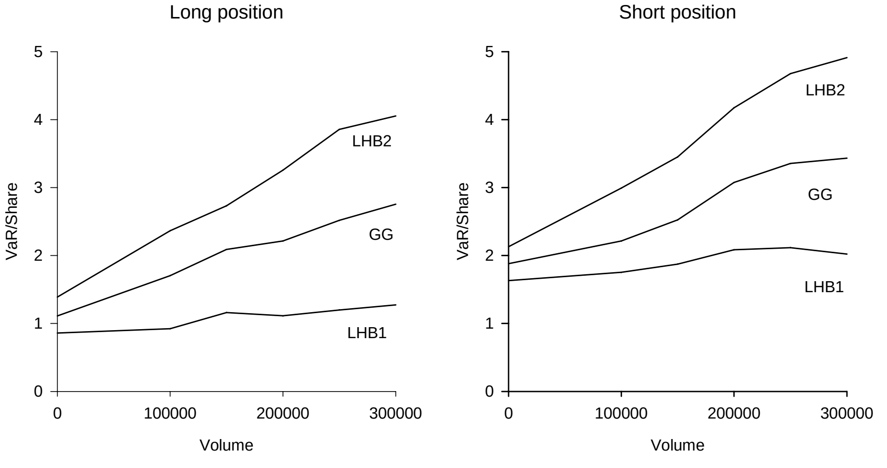

Figure 2: VaR per share vs volume for long and short positions in the SWB stock. GG refers to the VaR as given by the approach in Giot and Grammig (2006). LHB1 and LHB2 are our VaR's for a portfolio owned and purchased at T, respectively.

References

Bangia, A., Diebold, F. X., Schuermann, T., Stroughair, J. D., 1999. Modeling liquidity risk, with implications for traditional market risk measurement and management. Working paper 99-06, Wharton Financial Institutions Center.

Bowsher, C. G., 2004. Modelling the dynamics of cross-sectional price functions: An econometric analysis of the bid and ask curves of an automated exchange. Economics Papers 2004-W21, Economics Group, Nuffield College, University of Oxford.

Christoffersen, P., 2009. Value-at-risk models. In: Andersen, T. G., Davis, R. A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin Heidelberg, forthcoming.

Dionne, G., Duchesne, P., Pacurar, M., 2008. Intraday value at risk (IVaR) using tick-by-tick data with application to the Toronto stock exchange. Working paper, HEC Montréal.

Ernst, C., Stange, S., Kaserer, C., 2009. Measuring market liquidity risk - which model works best? CEFS Working paper series 1, Center for Entrepreneurial and Financial Studies, Technische Universität München.

François-Heude, A., Van Wynendaele, P., 2001. Integrating liquidity risk in a parametric intraday VaR framework. Working paper.

Giot, P., Grammig, J., 2006. How large is liquidity risk in an automated auction market. Empirical Economics 30, 867–887.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gourieroux, C., Jasiak, J., 2001. Financial Econometrics. Princeton University Press, Princeton, New Jersey.

Gourieroux, C., Le Fol, G., Meyer, B., 1998. Etude du carnet d'ordres. Banque et Marchés 36, 5–20.

Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill, New York.

Lönnbark, C., 2009. Uncertainty of multiple period risk predictors. Umeå Economic Studies 768, Department of Economics, Umeå University.

Malz, A., 2003. Liquidity risk: Current research and practise. RiskMetrics Journal 4, 35–72.